

METRON

RIVISTA INTERNAZIONALE DI STATISTICA — REVUE INTERNATIONALE DE STATISTIQUE
INTERNATIONAL REVIEW OF STATISTICS — INTERNATIONALE STATISTISCHE ZEITSCHRIFT
REVISTA INTERNACIONAL DE ESTADISTICA

DIRETTORE — DIRECTEUR — EDITOR — HERAUSGEBER — DIRECTOR

Dott. Corrado Gini, *Prof. emerito della Università di Roma*

COMITATO DIRETTIVO — COMITÉ DE DIRECTION — EDITORIAL COMMITTEE

DIREKTION-KOMITEE — COMITÉ DIRECTIVO

E.P. Billeter, Fribourg (Suisse); **F.P. Cantelli**, Roma (Italia); **V. Castellano**, Roma (Italia); **C. Dagum**, Cordoba (Argentina); **M.P. Geppert**, (Miss.), Frankfurt a. M. (Deutschland); **B. Gil**, Jerusalem (Israel); **L. Savage**, Chicago (U.S.A.); **H. von Schelling**, Schenectad New York (U.S.A.); **P. Steriotis**, Athens (Greece); **P. Whittle**, Wellington (New Zealand); **S. Wilks**, Princeton (U.S.A.); **H. Wold**, Stockholm (Sweden); **R. Yüceulug**, Ankara (Turkey).

SEGRETARI DI REDAZIONE — SECRÉTAIRES DE RÉDACTION

EDITORIAL SECRETARIES — REDAKTIONSSEKRETAERE — SECRETARIOS DE LA REDACCIÓN

Proff. C. Benedetti, T. Salvemini

Vol. XXI - N. 1-4

31-XII-1961

SOMMARIO-SOMMAIRE-CONTENTS-INHALT-SUMARIO

C. Gini e A. Trezza. <i>Anormalità delle distribuzioni statistiche dei caratteri antropologici</i>	pag. 3
Amato Herzel. <i>Sulla definizione dei concetti di media e di interpolazione</i>	» 126
Giuseppe Leti. <i>Nuovi tipi di distanze fra insiemi di punti e loro applicazioni alla statistica</i>	» 139
A.R. Kamat. <i>A note on Gini's mean difference</i>	» 170
Gerald J. Glasser. <i>Relationship between the mean difference and other measures of variation</i>	» 176
Carlo Benedetti. <i>A proposito dei rapporti tra differenza media e scostamenti medio quadratico, semplice medio e semplice medio dalla mediana</i>	» 181
Amato Herzel. <i>Le tabelle di co- e contrograduazione e la programmazione lineare</i>	» 186
Giuseppe Leti. <i>La distribuzione degli estremi e del campo di variazione dei campioni di una popolazione discreta e finita</i>	» 201
Alfredo Rizzi. <i>Sulle probabilità degli errori di compenso</i>	» 256
Amato Herzel. <i>Sul problema del « commesso viaggiatore »</i>	» 270
Corrado Gini. <i>Statistical Methods with special reference to Agriculture</i>	» 304
Lydia Spaventa De Novellis. <i>Evoluzione dell'indice di Tait attraverso il tempo</i>	» 393

DIREZIONE E REDAZIONE - DIRECTION ET REDACTION - EDITORIAL OFFICE -
DIREKTION UND REDAKTION - DIRECCIÓN Y REDACCIÓN

10, Via delle Terme di Diocleziano - Roma (ITALIA)

AMMINISTRAZIONE - ADMINISTRATEUR - MANAGER - VERWALTER-ADMINISTRACIÓN
Istituto di Statistica - Facoltà di Scienze Statistiche - Città Universitaria - Roma (ITALIA)

ARTICOLI GIUNTI ALLA RIVISTA CHE
VERRANNO PUBBLICATI NEI PROSSIMI
NUMERI (*secondo l'ordine di arrivo*).

ARTICLES RECEIVED BY THE REVIEW
WHICH WILL BE PUBLISHED IN FUTURE
ISSUES (*according to date of receipt*).

ARTIKEL DIE AN DIE ZEITSCHRIFT ANGEKOMMEN SIND UND WELCHE IN
DEN FOLGENDEN NUMMERN ERSCHEINEN WERDEN (*nach der Reihenfolge des
Eingangs*).

ARTICLES REÇUS PAR LA REVUE
ET À PARAÎTRE PROCHAINEMENT
(*d'après la date de réception*).

ARTICULOS LLEGADOS A LA REVISTA
QUE SE PUBLICARÁN EN LOS PROXIMOS
NUMEROS (*según el orden de su llegada*).

JOHN A. BRITAIN, *Interpolation of Frequency Distributions of
Aggregate Variables and Estimation of the Gini's Concentration
Measure.*

MAREK FISZ, *On an Inequality due to Gatti and Birnbaum.*

Gli Autori degli articoli inviati
per la pubblicazione nella Rivista
rinunciano in favore della medesima
alla proprietà letteraria degli articoli
stessi, qualora vengano pubblicati.

Les Auteurs des articles envoyés
à la Revue, pour y être publiés, renoncent, en faveur de celle-ci, à la
propriété littéraire de leurs articles,
s'ils sont acceptés.

Die Verfasser der zur Veröffentlichung in der Zeitschrift zugesandten Aufsätze, werden, falls selbige veröffentlicht werden, auf ihre Verfasserrrechte zu Gunsten der Zeitschrift verzichten müssen.

Los Autores de los artículos enviados para su publicación en « Metron » renuncian su propiedad a favor de la Revista cuando los artículos sean publicados.

The Authors of papers sent for publication in the Review are supposed to give up their copyright in favour of the Review if the papers are published.

C. GINI e A. TREZZA ⁽¹⁾

Anormalità delle distribuzioni statistiche dei caratteri antropologici

SOMMARIO. *Parte Prima*: 1. Osservazioni introduttive sulla teoria dell'ipernormalità e sulla sua misura e rappresentazione grafica. — 2-3. Dati su la statura, il perimetro toracico e il peso: 2 A) Dati per coetanei; 3 B) Dati relativi a popolazioni adulte. — 4. Dati per altri caratteri. — 5. Influenza dell'eredità e dell'ambiente sull'ipernormalità delle distribuzioni dei caratteri antropometrici. — 6. Riassunto della Parte Prima. — *Parte Seconda*: 7. Dati sulla rifrazione dell'occhio. — 8. Spiegazione della ipernormalità delle distribuzioni concernenti la rifrazione dell'occhio. — 8 bis. Spiegazione della loro asimmetria. — 9. Varie altre spiegazioni della ipernormalità delle distribuzioni dei caratteri antropometrici. — 10. Ipernormalità dei caratteri strutturali e dei caratteri funzionali. — *Parte Terza*: 11-12. Appendici teoriche; 11. Appendice Prima: Iponormalità attribuita alla composizione di due distribuzioni normali con uguale variabilità e media diversa. — 12. Appendice Seconda: Ipernormalità attribuita alla composizione di due distribuzioni aventi uguale media e diversa variabilità. 13. *Riassunto generale* (con traduzione in francese ed inglese).

PARTE PRIMA

OSSERVAZIONI INTRODUTTIVE SULLA TEORIA DELLA IPERNORMALITÀ E SULLA SUA MISURA E RAPPRESENTAZIONE GRAFICA

1. — Si definisce distribuzione normale una distribuzione che segue la legge $y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\delta}{\sigma}\right)^2}$, essendo σ il suo scarto quadratico medio e δ la media.

(1) La Dott. Anita Trezza, nella qualità di assistente straordinario alla cattedra di Statistica della Facoltà di Scienze statistiche, demografiche ed attuariali della Università di Roma, ha sorvegliato le elaborazioni degli indici di ipernormalità dal Prof. Gini affidate agli studenti universitari e ai calcolatori del Comitato Italiano per lo studio dei problemi della popolazione, dopo la pubblicazione della precedente memoria *La distribuzione delle stature è ipernormale*, e ha in gran parte steso, sotto la direzione del Prof. Gini, la Parte Prima di questo articolo. Ad essa sono anche dovute le *Appendici teoriche* che seguono il testo dell'articolo.

L'andamento di una curva normale o curva di Gauss è quello indicato in figura (1).

Si definisce invece *ipernormale* una distribuzione che ha una forma più snella della distribuzione normale di uguale variabilità,

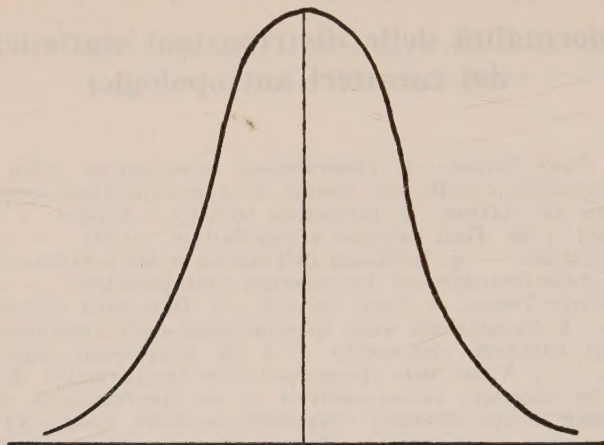


Fig. 1.

cioè tali che i valori centrali ed estremi sono in essa più frequenti. La curva risulta così più aguzza e più estesa di quella normale, come si vede in figura (2), dove la curva tratteggiata è quella normale.

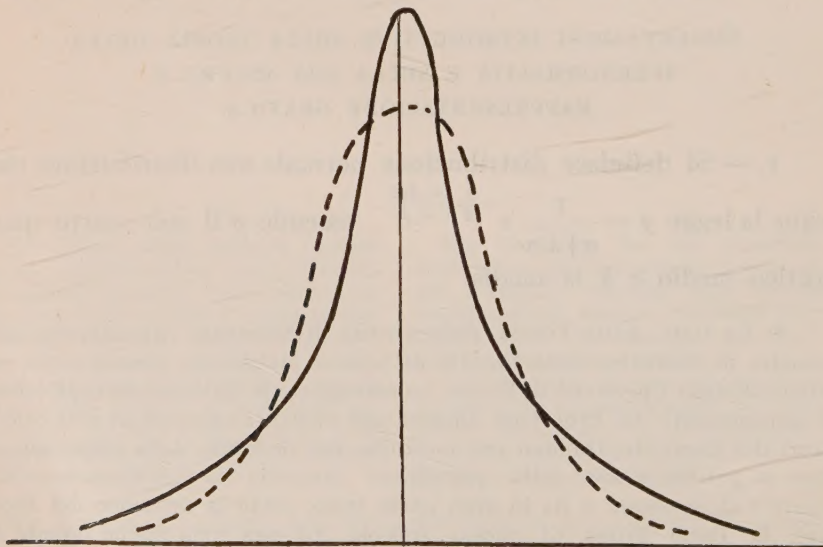


Fig. 2.

Si dice infine *iponormale* una distribuzione la cui forma è più tozza di quella normale avente la stessa variabilità, cioè tale che i valori centrali ed estremi sono in essa meno frequenti. La curva risulta così meno aguzza e meno estesa di quella normale, come si vede nella figura (3), dove pure la curva tratteggiata è quella normale.

Karl Pearson ha chiamato *curtosi* questo aspetto della conformazione delle curve, chiamando *platicurtiche* e *leptocurtiche* le curve, che noi abbiamo chiamato ipernormali e iponormali.

Per misurare la curtosi, il Pearson assume il rapporto tra il quarto momento e il quadrato del secondo momento intorno alla media aritmetica, cioè

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

Per una curva normale è $\beta_2 = 3$, per una ipernormale è $\beta_2 > 3$, per una curva iponormale è $\beta_2 < 3$.

Come è stato osservato dal Gini, il metodo del Pearson può venire generalizzato, prendendo come misura della ipernormalità la differenza tra il valore effettivo e quello teorico del rapporto tra due momenti di ordine diverso o tra due medie potenziate di diverso grado degli scostamenti. In questo metodo generalizzato rientra l'indice comunemente usato dal Gini, desunto dal quadrato del rapporto tra lo scarto quadratico medio e quello semplice. Dato che, in una distribuzione normale, il doppio del

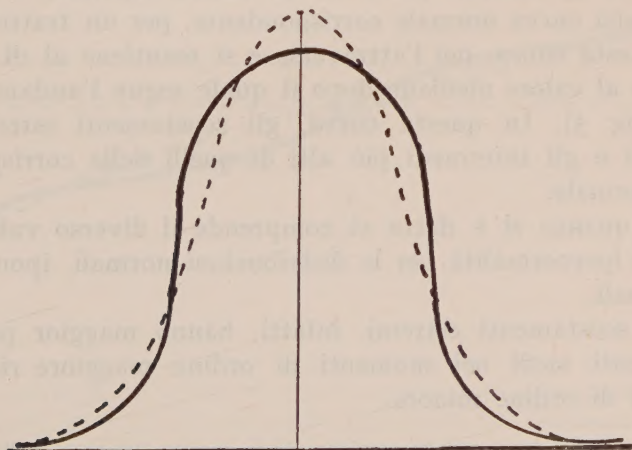


Fig. 3.

quadrato di questo rapporto è $\pi = 3,1416$, l'indice comunemente usato dal Gini è il seguente $I = 2 \frac{S^2}{1S^2}$. A seconda che la serie è ipernormale, normale o iponormale, si avrà:

$$I \begin{matrix} \geq \\ \leq \end{matrix} \pi$$

in cui, come è noto, $\pi = 3,1416$.

Questo indice è particolarmente adatto per il semplice controllo delle operazioni che il suo calcolo implica.

Un altro vantaggio di questo indice è quello di non dare troppo peso agli scostamenti estremi, fortemente influenzati del caso, rispetto a quelli intermedi.

È interessante esaminare le curve di graduazione corrispondenti ad una distribuzione simmetrica ipernormale e ad una simmetrica iponormale.

La prima (fig. 4) ⁽¹⁾ parte da un valore uguale o inferiore a quello della corrispondente curva normale, avente uguale media e variabilità, poi, dopo aver percorso un tratto al di sotto di questa curva, l'attraversa, mantenendosene al di sopra fino al valore mediano, dopo il quale segue l'andamento opposto. Gli scostamenti estremi sono più alti, mentre gli intermedi più bassi di quelli della corrispondente curva normale.

Viceversa, la curva di graduazione corrispondente ad una distribuzione iponormale parte da un valore superiore o uguale a quello della curva normale corrispondente, per un tratto decorre sopra questa curva, poi l'attraversa e si mantiene al di sotto di essa fino al valore mediano dopo il quale segue l'andamento opposto (fig. 5). In questa curva, gli scostamenti estremi sono più bassi e gli intermedi più alti di quelli della corrispondente curva normale.

Da quanto si è detto si comprende il diverso valore degli indici di ipernormalità per le distribuzioni normali, iponormali e ipernormali.

Gli scostamenti estremi, infatti, hanno maggior peso degli scostamenti medi nei momenti di ordine maggiore rispetto ai momenti di ordine minore.

⁽¹⁾ La curva tratteggiata corrisponde in questa figura e nella seguente alla distribuzione normale.

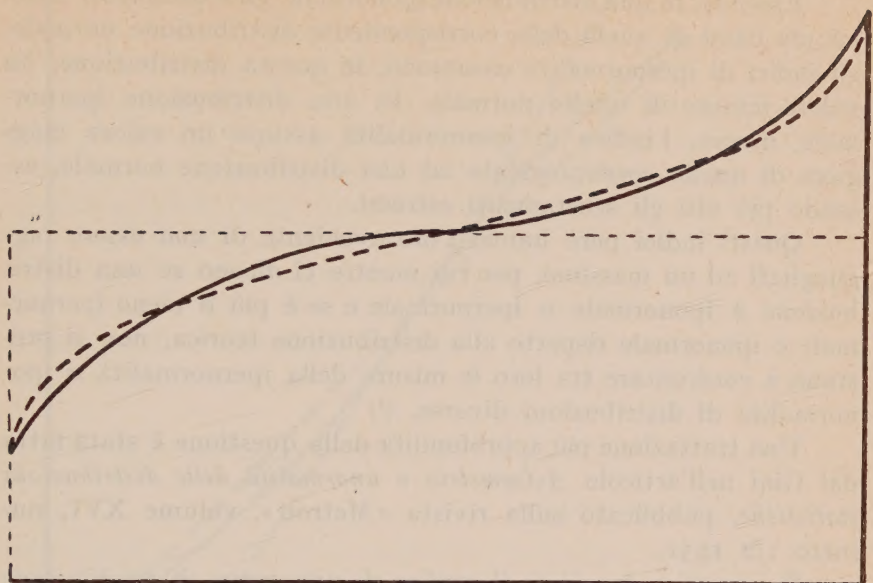


Fig. 4.

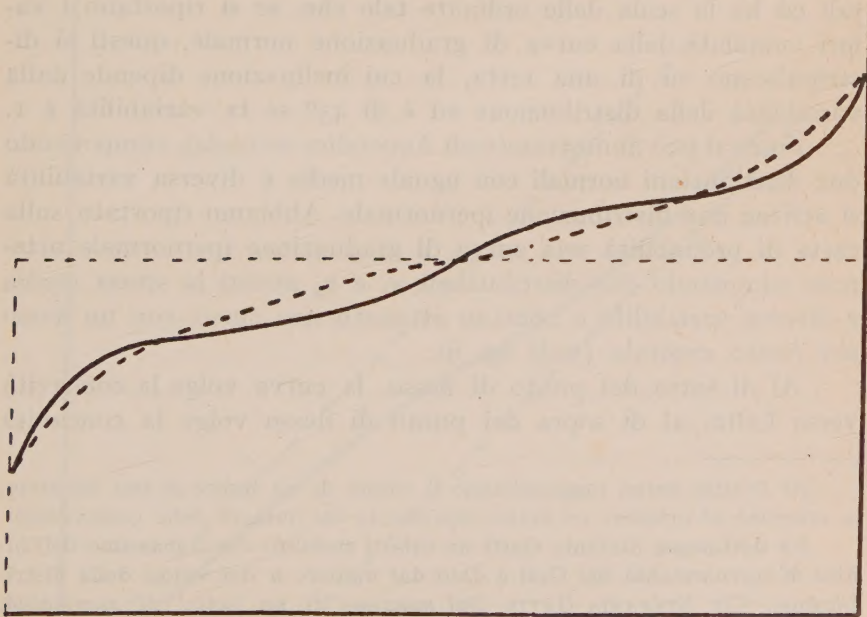


Fig. 5.

Essendo, in una distribuzione iponormale, gli scostamenti estremi più bassi di quelli della corrispondente distribuzione normale, gli indici di ipernormalità assumono, in questa distribuzione, un valore minore di quello normale. In una distribuzione ipernormale, invece, l'indice di ipernormalità assume un valore maggiore di quello corrispondente ad una distribuzione normale, essendo più alti gli scostamenti estremi.

Questi indici però hanno l'inconveniente di non essere ragguagliati ad un massimo, per cui mentre ci dicono se una distribuzione è iponormale o ipernormale e se è più o meno ipernormale o iponormale rispetto alla distribuzione teorica, non si prestano a confrontare tra loro le misure della ipernormalità o iponormalità di distribuzioni diverse. ⁽¹⁾

Una trattazione più approfondita della questione è stata fatta dal Gini nell'articolo *Asimmetria e anormalità delle distribuzioni statistiche*, pubblicato sulla rivista « Metron », volume XVI, numero 1/2 1951.

È istruttivo tracciare il grafico di una curva di graduazione ipernormale o iponormale sulla carta della probabilità. Come è noto, questa carta logaritmica porta sulle ascisse i valori naturali ed ha la scala delle ordinate tale che, se si riportano i valori cumulati della curva di graduazione normale, questi si distribuiscono su di una retta, la cui inclinazione dipende dalla variabilità della distribuzione ed è di 45° se la variabilità è 1.

Come si può dimostrare (vedi Appendice seconda), componendo due distribuzioni normali con uguale media e diversa variabilità si ottiene una distribuzione ipernormale. Abbiamo riportato sulla carta di probabilità una curva di graduazione ipernormale ottenuta sommando due distribuzioni y_1 e y_2 aventi la stessa media e diversa variabilità e abbiamo ottenuto una curva con un flesso nel punto centrale (vedi fig. 6).

Al di sotto del punto di flesso, la curva volge la concavità verso l'alto, al di sopra del punto di flesso volge la concavità

(1) D'altra parte, ragguagliando il valore di un indice al suo massimo, si verrebbe ad ottenere un indice dipendente dal numero delle osservazioni.

La dottoressa Stefania Gatti ha infatti stabilito che il massimo dell'indice di ipernormalità del Gini è dato dal numero n dei valori della distribuzione. (Cfr. STEFANIA GATTI, *Sul massimo di un indice di anormalità*, « Metron », vol. XVIII, n. 1-2, 1956).

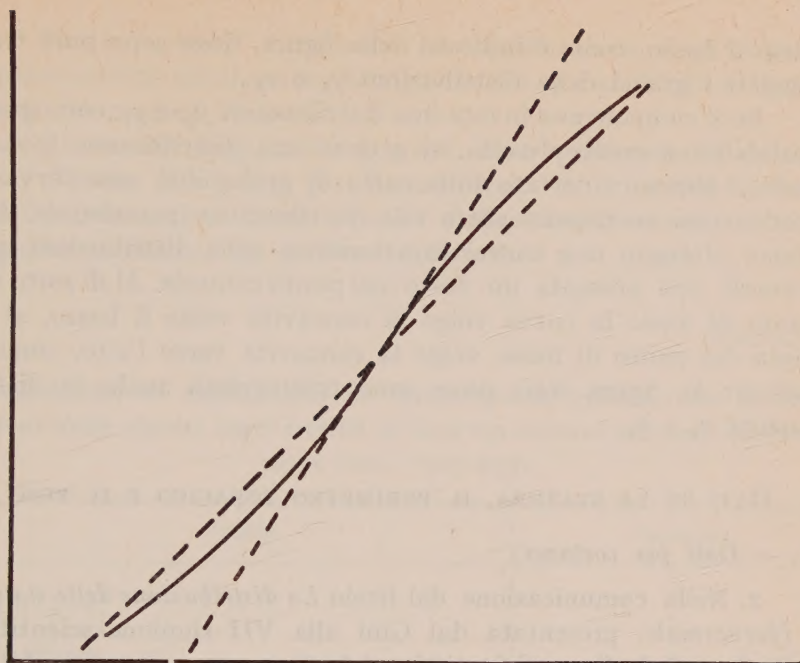


Fig. 6.

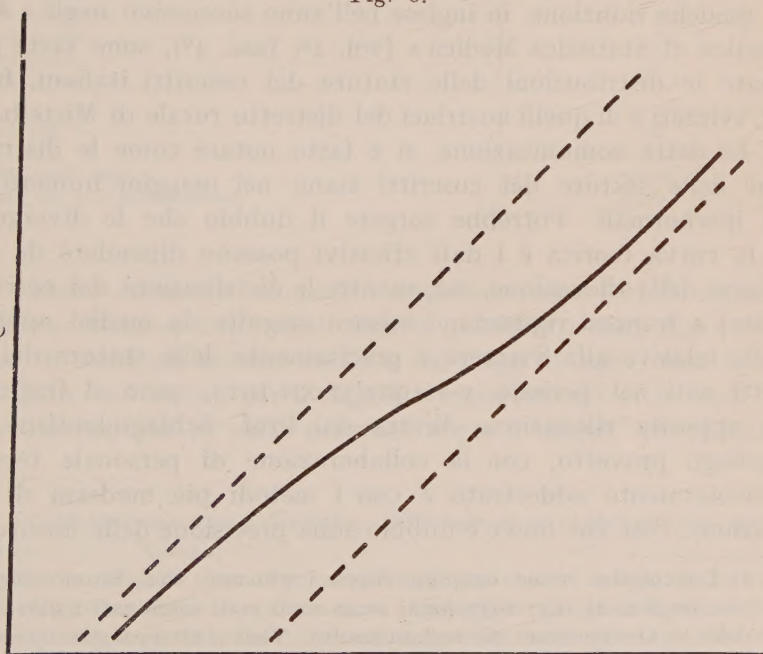


Fig. 6 bis.

verso il basso, come è indicato nella figura, dove sono pure tratteggiati i grafici delle distribuzioni y_1 e y_2 .

Se si compongono invece due distribuzioni y_1 , e y_2 , con uguale variabilità e media diversa, si ottiene una distribuzione iponormale. Abbiamo riportato sulla carta di probabilità una curva di graduazione corrispondente a tale distribuzione ipernormale. Abbiamo ottenuto una curva, caratteristica delle distribuzioni iponormali, che presenta un flesso nel punto centrale. Al di sotto del punto di flesso la curva volge la concavità verso il basso, al di sopra del punto di flesso volge la concavità verso l'alto, come è indicato in figura 6 bis dove sono tratteggiate anche le distribuzioni y_1 e y_2 .

DATI SU LA STATURA, IL PERIMETRO TORACICO E IL PESO

A. — *Dati per coetanei.*

2. Nella comunicazione dal titolo *La distribuzione delle stature è ipernormale*, presentata dal Gini alla VII riunione scientifica della Società Italiana di Statistica del 1° giugno 1949 e riprodotta, con qualche riduzione, in inglese nell'anno successivo negli « Acta Genetica et Statistica Medica » (vol. 1°, fasc. 4°), sono state esaminate le distribuzioni delle stature dei coscritti italiani, francesi, svizzeri e di quelli austriaci del distretto rurale di Mistelbach.

In detta comunicazione, si è fatto notare come le distribuzioni delle stature dei coscritti siano nel maggior numero dei casi ipernormali. Potrebbe sorgere il dubbio che le divergenze tra la curva teorica e i dati effettivi possano dipendere da inesattezze della rilevazione, ma, mentre le distribuzioni dei coscritti italiani e francesi riguardano misure eseguite da medici militari, quelle relative alla Svizzera e precisamente delle stature dei coscritti nati nel periodo 1-V-1907/31-XII-1913, sono il frutto di una apposita rilevazione diretta dal Prof. Schlaginhaufen, antropologo provetto, con la collaborazione di personale tecnico particolarmente addestrato e con i metodi più moderni di misurazione, così che non c'è dubbio sulla precisione delle misure (1).

(1) L'inchiesta venne eseguita sopra i giovani, che furono soggetti alla leva negli anni 1927-1932 (ogni anno sono stati esaminati i giovani di una delle 6 circoscrizioni di reclutamento). Vedi *Anthropologia Helvetica*: I. *Die Anthropologie der Eidgenossenschaft*, Zürich, Orell Füssli, 1946.

Nella tavola I riportiamo la classificazione degli indici di ipernormalità relativi alle distribuzioni di stature degli iscritti di leva nei cantoni svizzeri.

La media degli indici è 3.278. Essa praticamente coincide con l'indice (= 3.285) per il complesso dei coscritti di tutta la Confederazione. Come si vede dalla classificazione, soltanto quattro indici su 25 risultano inferiori a $\pi = 3,1416$.

TAVOLA I

Classificazione degli indici di ipernormalità relativi alle distribuzioni delle stature degli iscritti di leva nei cantoni svizzeri, esaminati negli anni 1927-1932

CLASSI	FREQUENZE
3,001 -3,100.	3
3,101 -3,1416	1
3,1416-3,200.	4
3,201 -3,300.	5
3,301 -3,400.	9
3,401 -3,500.	2
3,501 -3,600.	—
3,601 -3,700.	1
TOTALE.	25
Media degli indici	3 278
Indice per la Confederazione	3.285

Molto materiale su questo argomento ci è fornito dalle statistiche italiane. Sono stati infatti pubblicati dati sopra le stature di tutti gli iscritti di leva dal 1854 al 1920, non solo per lo Stato, ma anche per i singoli compartimenti e, in ogni compartimento, per le varie province e per i singoli circondari. ⁽¹⁾

⁽¹⁾ MINISTERO DELLA GUERRA, *Delle leve di terra sui giovani nati nell'anno*. Roma.

Per le classi di leva anteriori al 1854, con cui si iniziò la pubblicazione della Relazione sulle leve di terra, si posseggono alcuni dati, che il Bodio, allora professore a Venezia, comunicò al Quetelet e che questi pubblicò nella sua *Fisica sociale* (Vedi traduzione italiana in « Biblioteca dell'Economista »,

La tavola seguente fornisce i valori dell'indice di ipernormalità per gli iscritti di leva dei dieci anni 1854, 1860, 1870, 1880, 1890, 1900, 1910, 1915, 1918, 1920, sia per i singoli compartimenti, sia per tutto lo Stato.

TAVOLA II

Indici di ipernormalità delle distribuzioni relative alla statura degli iscritti di leva nei compartimenti italiani.

COMPARTIMENTI	ANNI DI NASCITA									
	1854	1860	1870	1880	1890	1900	1910	1915	1918	1920
Piemonte	3.484	3.590	3.353	3.335	3.321	3.375	3.304	3.232	3.243	3.258
Liguria	3.321	3.286	3.254	3.363	3.271	3.275	3.242	3.301	3.253	3.274
Lombardia	3.457	3.521	3.410	3.298	3.328	3.326	3.233	3.268	3.248	3.247
Veneto	3.295	3.287	3.297	3.238	3.332	3.354	3.231	3.228	3.226	3.270
Emilia	3.263	3.257	3.262	3.295	3.283	3.331	3.218	3.260	3.235	3.228
Toscana	3.249	3.260	3.261	3.218	3.146	3.304	3.261	3.193	3.241	3.240
Marche	3.280	3.243	3.228	3.230	3.236	3.344	3.250	3.251	3.230	3.261
Umbria	3.264	3.310	3.239	3.275	3.452	3.359	3.213	3.302	3.174	3.226
Lazio	3.312	3.362	3.446	3.429	3.346	3.351	3.191	3.292	3.241	3.064
Abruzzi e Molise	3.257	3.366	3.363	3.347	3.238	3.264	3.291	3.161	3.197	3.206
Campania	3.300	3.294	3.355	3.294	3.294	3.276	3.162	3.198	3.356	3.330
Puglie	3.483	3.335	3.370	3.472	3.301	3.316	3.343	3.318	3.236	3.198
Basilicata	3.367	3.419	3.532	3.376	3.280	3.292	3.263	3.305	3.249	3.382
Calabria	3.429	3.459	3.562	3.368	3.354	3.294	3.233	3.216	3.289	3.182
Sicilia	3.336	3.415	3.390	3.357	3.430	3.386	3.238	3.254	3.314	3.295
Sardegna	3.396	3.413	3.408	3.333	3.466	3.434	3.291	3.196	3.208	3.301
Venezia Tridentina	—	—	—	—	—	—	—	3.245	3.406	3.433
Venezia Giulia e Zara	—	—	—	—	—	—	—	3.274	3.260	3.189
ITALIA	3.350	3.378	3.344	3.320	3.288	3.339	3.243	3.224	3.248	3.231

Dalla tavola si desume che, su 166 indici relativi ai compartimenti, soltanto uno (e precisamente quello relativo alla statura degli iscritti di leva del Lazio nel 1920) risulta inferiore a

serie III, vol. II, Unione tipografico-editrice, Torino, 1878, pag. 649). Essi si riferiscono ai giovani nati negli anni 1943-44-45 (leve anticipate). Il Bodio informa di averli desunti dalle pubblicazioni ufficiali del Ministero della Guerra. In contrasto coi risultati ottenuti per le serie degli anni posteriori, essi mostrerebbero una distribuzione lievemente iponormale ($I = 3,135$). È possibile che ciò dipenda, in tutto o in parte, dal fatto che erano stati trascurati, come il Bodio informa, gli individui con statura inferiore a m. 1,33.

3,1416, mentre, degli indici relativi all'Italia, nessuno è inferiore a tale valore. La media degli indici dei compartimenti è 3,322, mentre la media degli indici relativi all'Italia è 3,297.

Dai dati dell'ultima linea si desume pure che, attraverso il periodo considerato (66 anni), gli indici di ipernormalità, sia pure oscillando, hanno mostrato tendenza a diminuire ⁽¹⁾.

La tavola III mostra come si distribuiscano i valori trovati per i singoli compartimenti. Essi presentano la massima frequenza nella classe da 3,201 a 3,299.

TAVOLA III

*Seriazione di 166 indici di ipernormalità delle stature
nei compartimenti italiani.*

CLASSI	FREQUENZE
3.000 -3.100.	1
3.100 -3.1416	—
3.1417-3.200.	12
3.200 -3.300.	81
3.300 -3.400.	49
3.400 -3.500.	19
3.500 -3.600.	4
TOTALE.	166
Media dei 166 indici per compartimento	3.322
Media dei 10 indici per lo Stato	3.297

La tavola IV fornisce gli indici di ipernormalità nelle province lombarde e toscane nel 1915. Su 18 indici di ipernormalità, uno solo, e precisamente quello relativo a Livorno (con 3,118), è inferiore a 3,1416 e la media degli indici è, per le province lombarde, 3,242 e, per le province toscane, 3,216.

⁽¹⁾ Questa tendenza è confermata e resa più evidente dal calcolo degli indici di ipernormalità che, col metodo del Pearson, ha eseguito, per i singoli anni dal 1874 al 1916, il prof. LUIGI FALESCHINI in una memoria *Sui miscugli di distribuzioni gaussiane*, pubblicata in « Rivista Italiana di Economia Demografica e Statistica », fascicolo 1^o, 1951.

La tavola V ci dà i valori degli indici di ipernormalità delle distribuzioni delle stature dei coscritti delle province e dei circondari della Lombardia nel 1898. Su 25 indici relativi ai circondari soltanto tre (Chiari, Salò, Varese) sono inferiori a 3,1416 e la media dei 25 indici è 3,268. Tutti gli indici relativi alle province sono invece superiori a 3,1416, con una media di 3,280.

La tavola VI fornisce i valori degli indici di ipernormalità delle province e dei circondari della Toscana nel 1898. Su 16 indici relativi ai circondari, soltanto due (Isola d'Elba, Lucca) sono inferiori a 3,1416. La media dei 16 indici è 3,292. Gli indici relativi alle province sono tutti superiori a 3,1416, con una media di 3,285.

TAVOLA IV

Indici di ipernormalità delle stature nelle province lombarde e toscane (1915).

LOMBARDIA		TOSCANA	
Province	Indici	Province	Indici
Bergamo.	3.262	Arezzo.	3.226
Brescia	3.168	Firenze	3.206
Como	3.234	Grosseto.	3.186
Cremona.	3.286	Livorno	3.118
Mantova.	3.235	Lucca	3.240
Milano.	3.238	Massa Carrara	3.258
Pavia	3.245	Pisa.	3.163
Sondrio	3.401	Pistoia.	3.329
Varese.	3.190	Siena	3.218
MEDIA DEGLI INDICI . . .	3.242	MEDIA DEGLI INDICI . . .	3.216
INDICE DEL COMPARTIMENTO	3.268	INDICE DEL COMPARTIMENTO	3.193

Da tutto ciò che abbiamo visto risulta accertato il fatto che le distribuzioni delle stature dei coscritti in Italia, almeno negli anni considerati, sono nettamente ipernormali.

Si può notare inoltre che gli indici di ipernormalità per le circoscrizioni più vaste (Stato, compartimenti, province) non appaiono superiori a quelli per le circoscrizioni meno vaste (rispettivamente, compartimenti, province, circondari), ciò che suggerisce la conclusione che la ipernormalità non è correlata colla eterogeneità della popolazione.

TAVOLA V

Indici di ipernormalità delle stature nelle province e nei circondari della Lombardia.

(1898)

PROVINCE	INDICI	CIRCONDARI	INDICI
Bergamo	3.283	{ Bergamo	3.277
		{ Clusone.	3.349
		{ Treviglio	3.271
Brescia	3.241	{ Breno	3.335
		{ Brescia.	3.274
		{ Chiari	3.027
		{ Salò	3.128
		{ Verolanuova	3.327
Como	3.374	{ Como	3.384
		{ Lecco	3.374
		{ Varese	3.121
Cremona	3.277	{ Casalmaggiore.	3.330
		{ Crema	3.210
		{ Cremona	3.285
Mantova	3.163	Mantova	3.163
Milano	3.282	{ Abbiategrasso.	3.309
		{ Gallarate	3.204
		{ Lodi	3.243
		{ Milano	3.257
		{ Monza	3.371
Pavia	3.325	{ Bobbio.	3.221
		{ Lomellina.	3.221
		{ Pavia	3.341
		{ Voghera	3.428
Sondrio.	3.292	Sondrio.	3.292
MEDIA DEGLI INDICI .	3.280	MEDIE DEGLI INDICI .	3.268

INDICE DEL COMPARTIMENTO = 3,297

Per quanto riguarda il confronto tra lo Stato e i compartimenti, questa conclusione si desume, oltre che dalla tav. III, dalla tav. VII, nella quale vengono riportati gli indici di ipernormalità per l'Italia e le medie semplici e ponderate (secondo il numero dei coscritti) degli indici di ipernormalità dei compar-

Indici di ipernormalità delle stature nelle province e nei circondari della Toscana.

(1898)

PROVINCE	INDICI	CIRCONDARI	INDICI
Arezzo	3.515	Arezzo	3.515
Firenze.	3.370	{ Firenze.	3.433
		{ Pistoia	3.224
		{ Rocca S. Casciano.	3.274
		{ S. Miniato	3.266
Grosseto	3.261	Grosseto	3.261
Livorno	3.250	{ Isola d'Elba	3.097
		{ Livorno	3.292
Lucca	3.076	Lucca	3.076
Massa e Carrara.	3.420	{ Castelnovo.	3.219
		{ Massa	3.395
		{ Pontremoli	3.523
Pisa	3.181	{ Pisa	3.169
		{ Volterra	3.314
Siena.	3.203	{ Montepulciano	3.329
		{ Siena.	3.192
MEDIA DEGLI INDICI .	3.285	MEDIE DEGLI INDICI .	3.286

INDICE DEL COMPARTIMENTO = 3.299

timenti. Si nota infatti come queste ultime risultino per cinque anni superiori e per quattro inferiori all'indice per l'Italia (e sempre con lievi differenze), se ci si basa sulla media semplice, e per sei anni superiori, per due inferiori e per uno eguali, se ci si basa sulla media ponderata.

D'altra parte, dalla tavola IV si rileva che nel 1915 l'indice per la Lombardia (3,268) risulta lievemente superiore alla media degli indici per le sue province (3,242) e viceversa l'indice per la Toscana (3,193) lievemente inferiore alle medie degli indici per le sue province (3,216), mentre dalle tavole V e VI risulta

TAVOLA VII

Indici di ipernormalità delle stature per l'Italia.

A N N I	MEDIE DEGLI INDICI DI IPERNORMALITÀ DEI COMPARTIMENTI		INDICE DI IPERNORMALITÀ DELL'ITALIA
	semplici	ponderate	
1854	3.343	3.356	3.350
1860	3.364	3.378	3.378
1870	3.358	3.356	3.344
1880	3.327	3.317	3.320
1890	3.317	3.313	3.288
1900	3.330	3.332	3.339
1915	3.250	3.245	3.224
1918	3.256	3.257	3.248
1920	3.225	3.246	3.231
MEDIE	3.308	3.311	3.302

che nel 1898 la media degli indici delle province lombarde (3,280) risulta lievemente superiore alle medie degli indici dei rispettivi circondari (3,268) e le medie degli indici delle province toscane (3,285) appena inferiore alle medie degli indici dei rispettivi circondari (3,286). Queste medie risultano solo lievemente inferiori agli indici per i rispettivi compartimenti (per la Lombardia = 3,297; per la Toscana = 3,299). La nostra conclusione appare quindi confermata anche dai confronti tra gli indici di ipernormalità dei compartimenti e delle province, nonchè tra gli indici di ipernormalità delle province e dei rispettivi circondari.

Ulteriore conferma si trae dal confronto tra l'indice di ipernormalità trovato per la Confederazione svizzera (= 3.285) e la media degli indici per i 25 cantoni (= 3.278) che praticamente coincidono (Cfr. pag. 11).

Classificazioni analoghe a quelle dell'Italia contengono i resoconti delle leve pubblicati dalla Francia per gli iscritti di leva nati tra il 1903 al 1922 (mancano i dati per il 1913 e il 1915) ⁽¹⁾.

⁽¹⁾ *Compte Rendu sur le Recrutement de l'Armée pendant l'Année...* Paris, Imprimerie Nationale.

Anche in questo caso, l'indice risulta sempre superiore al valore teorico e varia tra 3,20 e 3,32 con una media di 3,250. Nella tavola VIII sono riportati gli indici di ipernormalità per i 18 anni considerati e le medie semplici e ponderate (secondo il numero dei coscritti) degli indici annuali.

Notiamo che queste medie non risultano inferiori all'indice globale, e da questo deduciamo che, scindendo le distribuzioni nel tempo, non si ha riduzione della ipernormalità.

TAVOLA VIII

Indici di ipernormalità per le stature degli iscritti di leva francesi.

ANNI DI NASCITA		INDICI
1903		3.212
1904		3.225
1905		3.215
1906		3.201
1908		3.222
1909		3.246
1910		3.242
1911		3.245
1912		3.226
1914		3.249
1916		3.261
1917		3.324
1918		3.285
1919		3.293
1920		3.317
1921		3.268
1922		3.217
Periodo intero.		3.244
MEDIA DEGLI INDICI ANNUALI	semplce.	3.250
	ponderata	3.249

Notiamo anche che, a differenza di quanto abbiamo constatato per l'Italia, la serie degli indici non mostra alcuna tendenza a diminuire col tempo.

Altri dati per la popolazione francese sono contenuti in un articolo recente. Essi si riferiscono alla statura degli studenti dell'insegnamento superiore, in base alle misure effettuate in occasione dell'esame, a cui essi sono sottoposti sistematicamente ogni anno da parte della « Medicina Preventiva Universitaria ». Seriazioni delle stature sono state pubblicate per gli studenti in età dai 20 ai 24 anni, misurati nel 1947 e nel 1954. ⁽¹⁾ Gli indici di ipernormalità risultano rispettivamente $= 3,212$ e $= 3,206$, ⁽²⁾ valori sostanzialmente concordanti con quelli della tavola VIII relativi ai coscritti, anche se lievemente inferiori alle medie di questi.

Dal trattato del Winkler ⁽³⁾ sono stati elaborati dati relativi alla distribuzione delle stature dei coscritti del distretto rurale di Mistelbach, nell'Austria inferiore, per l'anno 1913. Anche in questo caso, l'indice risulta superiore a 3,1416 e precisamente $= 3,204$.

Dati che credo inediti, relativi alla statura di 71.664 coscritti belgi esaminati nel 1931, sono esposti nella tavola VIII *bis*. Per essi, l'indice di ipernormalità ($= 3,165$) risulta di poco superiore al valore normale (3,1416).

Per i coscritti belgi, il Quetelet aveva già pubblicato nella *Fisica sociale* dati sulla statura molto numerosi (911.087 casi)

⁽¹⁾ Vedi: M. AUBENQUE - *Note documentaire sur la Statistique des tailles des étudiants au cours de ces dernières années*, in « Biotypologie », tome XVIII, n. 3, settembre 1957, pag. 211.

⁽²⁾ Questi valori degli indici di ipernormalità sono stati ottenuti attribuendo il valore di 145 alle stature della prima classe con stature inferiori a 150 cm. Se invece si attribuisce ad esse il valore di 148, si ottiene, sia per il 1947 che per il 1954, un indice d'ipernormalità $1 = 3,200$. Questi risultati mostrano come, quando le classi estreme sono aperte, esiste un certo grado di incertezza nella determinazione dell'indice di ipernormalità a seconda che, per le unità che cadono in dette classi, si assuma un valore piuttosto che un altro. Quando qualcuna delle serie per cui si mettono a confronto gli indici di ipernormalità possiede delle classi aperte, è dunque necessario non attribuire importanza a piccole differenze che risultino negli indici di ipernormalità.

⁽³⁾ *Grundriss der Statistik: I, Theoretische Statistik*, Erste Auflage, Springer, 1931, pp. 21-22; zweite Auflage, Wien, Manzsche Verlag, 1947, pag. 35.

TAVOLA VIII-bis
Distribuzione degli iscritti di leva in generale e degli abili in particolare in Belgio (visitati nel 1931).

STATURA				PERIMETRO TORACICO				PESO			
Centimetri	Abili	Abili + inabili	Frequenze degli inabili (°/oo)	Centimetri	Abili	Abili + inabili	Frequenze degli inabili (°/oo)	Chilogrammi	Abili	Abili + inabili	Frequenze degli inabili (°/oo)
Meno 151	5	431	988	Meno 70.	16	171	905	Meno 40.	1	152	993
151-153	238	573	585	70-72	7	284	975	40-42	—	165	1,000
153-155	476	836	431	72-74	42	970	957	42-44	12	395	970
155-157	939	1,511	379	74-76	218	2,462	911	44-46	75	875	914
157-159	1,693	2,604	350	76-78	1,427	5,286	730	46-48	348	1,679	793
159-161	3,005	4,427	321	78-80	4,932	9,047	455	48-50	1,059	3,113	660
161-163	4,344	6,315	312	80-82	9,034	12,231	261	50-52	2,308	4,705	509
163-165	5,654	8,165	308	82-84	10,311	12,932	204	52-54	3,602	6,269	425
165-167	6,323	9,056	302	84-86	9,234	11,281	181	54-56	5,142	7,675	330
167-169	6,445	9,159	296	86-88	6,219	7,610	183	56-58	6,221	8,492	267
169-171	5,977	8,555	301	88-90	3,821	4,661	180	58-60	6,761	8,737	326
171-173	4,866	6,930	298	90-92	1,948	2,355	173	60-62	5,886	7,438	209
173-175	3,769	5,311	290	92-94	984	1,222	195	62-64	5,038	6,266	196
175-177	2,375	3,428	307	94-96	431	527	182	64-66	3,818	4,760	198
177-179	1,401	2,066	322	96-98	210	262	198	66-68	2,972	3,643	184
179-181	824	1,225	327	98-100	102	130	215	68-70	2,123	2,638	195
181-183	390	580	328	oltre 100	118	213	446	70-72	1,273	1,568	188
183-185	183	279	344		—	—	—	72-74	857	1,061	192
oltre 185	147	213	310		—	—	—	74-76	532	648	179
	—	—	—		—	—	—	76-78	333	445	252
	—	—	—		—	—	—	78-80	239	302	209
	—	—	—		—	—	—	oltre 80	454	638	288
	49,054	71,664	316		49,054	71,664	316		49,054	71,664	316

relativi alle classi di leva dal 1842 al 1865. Senonchè la categoria di statura più bassa (sotto m. 1.560) è troppo ampia (comprendendo quasi il 14 % dei casi) perchè il calcolo di un indice di ipernormalità risulti attendibile.

Ad illustrare la ipernormalità della distribuzione della statura degli iscritti di leva può ben servire il confronto, con le teoriche, delle frequenze effettive degli scostamenti di varia intensità che si presentano nelle distribuzioni. Per le distribuzioni relative all'Austria (Mistelbach), Belgio, Francia, Italia e Svizzera il confronto è eseguito nella tavola VIII *ter*. Come da essa

TAVOLA VIII-*ter*

Confronto della distribuzione effettiva con la distribuzione teorica degli iscritti di leva in vari Stati (Austria, Belgio, Francia, Italia, Svizzera)

INTENSITÀ DELLO SCOSTAMENTO RAGGUAGLIATO ALLO SCOSTAMENTO QUADRATICO MEDIO	FREQUENZA CON CUI UNO SCOSTAMENTO PRESENTA L'INTENSITÀ INDICATA ALLA COL. I					
	Frequenza teorica in base alla curva normale	Frequenza effettiva (1) degli iscritti di leva				
		Austria	Belgio	Francia	Italia	Svizzera
		Mistel- bach nati nel 1913	Reclute visitate nel 1931	nati 1903- 1922	nati 1900	nati 1907- 1913
0 -1/4 . . .	19,74	20,68	19,86	20,72	21,09	20,56
1/4-1/2 . . .	18,56	19,37	18,74	18,88	19,99	19,27
1/2-3/4 . . .	16,38	15,37	16,57	16,69	16,84	16,96
3/4-1 . . .	13,58	12,64	13,67	14,08	12,73	13,33
1 -1 1/2 . . .	18,38	17,86	17,92	16,93	16,66	17,23
1 1/2-2 . . .	8,80	12,57	8,26	7,88	7,70	8,06
2 -2 1/2 . . .	3,32		3,36	2,88	3,11	2,92
2 1/2-3 . . .	0,97	1,51	1,32	1,31	1,12	1,04
3 -3 1/2 . . .	0,22		0,40	0,48	0,38	
3 1/2-4 . . .	0,04		0,30	0,23	0,12	
più di 4 . . .	0,01	0,05	0,27	1,24	0,13	
	100.—	100 —	100.—	100.—	100.—	100.—

(1) Sono in corsivo le frequenze effettive che risultano superiori alle teoriche.

si vede, la frequenza effettiva risulta, in tutte le serie, superiore alla teorica per le classi degli scostamenti più piccoli e più forti, mentre risulta inferiore per gli scostamenti intermedi, come è caratteristica delle curve ipernormali.

Interesserebbe l'esame delle distribuzioni delle stature dei coscritti presentatisi o arruolati nelle armi scelte per le quali generalmente si richiedono speciali requisiti fisici.

Dati sono stati pubblicati in Italia sulle stature di giovani italiani presentatisi in vista dell'arruolamento volontario nella marina ⁽¹⁾. In questo caso, sono stati distinti i giovani secondo l'età.

Si sono ottenuti gli indici d'ipernormalità riportati nella tavola IX.

Si può notare come la distribuzione relativa ai giovani di 15 anni presenti una forte iponormalità, che molto probabilmente deriva dal fatto che una parte dei giovani di quella età non ha ancora completato lo sviluppo e quindi la curva risulta dalla somma

TAVOLA IX

Indici di ipernormalità per le stature
(Dati degli « Annali di medicina navale »).

ANNI DI ETÀ	INDICE DI IPERNORMALITÀ
15 (abili più inabili)	2.956
16 (abili più inabili)	3.210
17 (abili più inabili)	3.245
18 (abili più inabili)	3.158
19 (abili più inabili)	3.341
15-19 (abili più inabili)	3.226
16-19 (abili più inabili)	3.339
Media degli indici relativi alle età 15-16-17-18-19	3.182
Media degli indici relativi alle età 16-17-18-19	3.238

⁽¹⁾ L. SESTINI - *Sui criteri desunti dai dati fisici per determinare la attitudine al servizio militare nell'armata italiana*. (Estratto dagli « Annali di Medicina Navale » Anno XI, vol. II, fasc. V, Novembre 1905).

di più curve con medie diverse. Le altre distribuzioni sono tutte ipernormali. Sono stati anche calcolati gli indici medi e gli indici relativi alla distribuzione complessiva e non si avverte una differenza notevole del valore dell'indice tra le distribuzioni parziali e la complessiva. Se si paragonano gli indici delle cinque distribuzioni parziali con l'indice relativo ai giovani dai 15 ai 19 anni, si nota che tre indici sono inferiori a quello relativo alla distribuzione complessiva, mentre, se si paragonano gli ultimi quattro indici delle distribuzioni parziali con quello relativo alla distribuzione dei giovani dai 16 ai 19 anni, che vale 3,339, uno di questi indici supera quello relativo alla distribuzione complessiva, e precisamente quello relativo ai giovani di 19 anni (che vale 3,341), mentre gli altri ne sono alquanto inferiori.

L'arruolamento implica una selezione più o meno severa a seconda delle armi a cui le reclute sono destinate e dei criteri in essa adottati.

Dei 71.664 coscritti belgi dell'anno 1931, di cui sopra abbiamo dato l'indice di ipernormalità per le stature (pag. 19), 49.054 furono dichiarati abili. La distribuzione degli uni e degli altri è data alla tavola VIII *bis*. Per gli abili l'indice di ipernormalità da 3,165 discende a 3,138 (tavole IX-*ter*), risultando così inferiore al valore (3,1416) corrispondente alla distribuzione normale.

Anche due serie di stature di militari americani riportate dal Quetelet ⁽¹⁾ danno indici iponormali (3,094 e 3,021).

Nella tavola IX-*bis* sono riportati, oltre a questi, altri indici per serie di stature di militari in servizio.

Il Gini aveva infatti trovato un risultato analogo (indice = 3,094) per i soldati italiani nati negli anni 1859-63, mentre, per i militari svedesi, l'indice risultava = 3,159, nell'inchiesta eseguita nel 1897-98 da Retzius e Fürst, ed = 3,12 nell'inchiesta eseguita nell'anno 1922-24 da Lundborg e Linders ⁽²⁾. Si può dun-

⁽¹⁾ Vedi *Antropometria*, traduzione italiana in « Biblioteca dell'Economista », serie III, vol. II, a pag. 1131 per i dati dei 25878 volontari americani e a pag. 1149 per quelli dei 761 soldati diversi, pure americani, secondo B.A. Gould.

⁽²⁾ C. GINI, *La distribuzione delle stature è ipernormale*, « Atti dell'VIII riunione scientifica della Società italiana di Statistica », 1-2 Giugno 1949, nota a pag. 34.

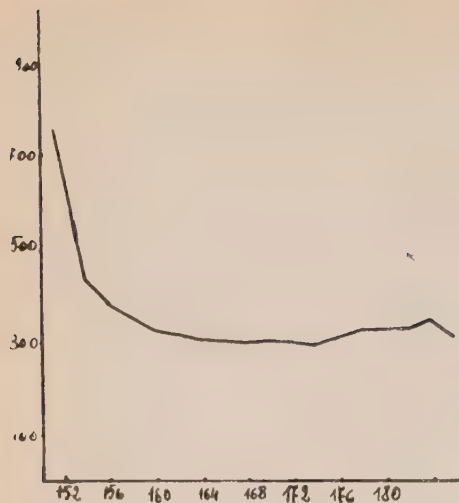


Fig. 7.

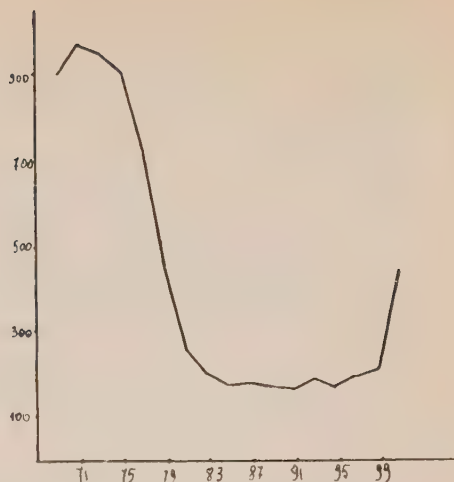


Fig. 7 bis.

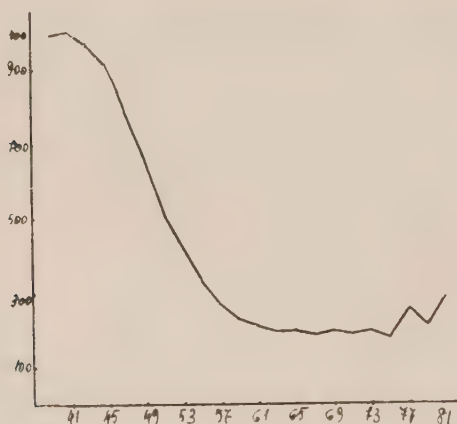


Fig. 7 ter.

que dire che, per i militari, l'indice risulta più basso che per gli iscritti di leva e spesso al disotto del valore normale. Ciò può attribuirsi al fatto che, all'atto del reclutamento, c'è una selezione per cui vengono eliminate con particolare frequenza sia le stature basse sia anche quelle molto alte. Nella tavola VIII-bis sono indicate le percentuali dei dichiarati inabili secondo la sta-

TAVOLA IX-bis

Indici di ipernormalità per le stature, i perimetri toracici e i pesi di militari in servizio.

	STATURE	PERIMETRI TORACICI	PESI
U.S.A. - 25.878 volontari della guerra di secessione.	3.094		
716 soldati diversi.	3.021		
Armata del Potomac (1716 soldati) . .		3.322	
SCOZIA - 5738 soldati		3.157	
ITALIA - Soldati dell'esercito: classi di leva 1859-64.	3.09		
SVEZIA - Inchiesta di Retzius e Fürst . .	3.16		
Inchiesta di Lundborg e Linders (1922-24)	3.12		
FINLANDIA - 2.600 reclute degli anni 1901-1902-1903.	3.126	3.310	3.216

tura e i diagrammi qui sopra riportati (figg. 7, 7-bis, 7-ter) mettono in evidenza la eliminazione più forte a cui sono sottoposte le stature basse e, sia pure in misura molto minore, anche le più elevate (1). È pure intuitivo che l'influenza della sele-

TAVOLA IX-ter

Indici di ipernormalità dei coscritti (abili + inabili) belgi visitati nel 1931 a confronto con quelli dei dichiarati abili.

	STATURE	PERIMETRI TORACICI	PESI
Abili + Inabili.	3.165	3.412	3.355
Abili	3.138	3.244	3.276

(1) Ai fini della rappresentazione, sono state raggruppate le due classi più basse delle stature, sotto 153 cm.: 243 coscritti di leva risultarono abili e 1034 inabili, con una percentuale di inabili sul totale di 765 per mille.

zione risulta minore tra popolazioni più alte, come la svedese, che tra popolazioni più basse, come la italiana.

Altri dati relativi alle stature di soldati abbiamo ricavati, da rilevazioni effettuate alla fine del 1700 e al principio del 1800, su soldati finnici ⁽¹⁾. Abbiamo calcolato l'indice di ipernormalità per una distribuzione relativa alle stature di 18.406 soldati appartenenti a tutto il Paese ed abbiamo ottenuto per esso il valore 3,405. Abbiamo poi studiato altre due distribuzioni, delle quali la prima si riferisce a soldati svedesi appartenenti alle tre province (Landschaften) di Varsinaissuomi, Uusimaa, Pohjanmaa, e la seconda ai soldati finnici appartenenti alle stesse province, ottenendo rispettivamente i valori 3,899 e 3,479.

Come si vede, questi tre indici sono molto alti e ciò può, con molta verosimiglianza, mettersi in relazione col fatto che, come viene dichiarato nell'articolo da cui questi dati sono stati tolti, la visita medica non era allora rigorosa, così che venivano ammessi nell'esercito anche individui che oggi comunemente verrebbero scartati. Si deve notare però che la ipernormalità risulta particolarmente forte, più forte, cioè, che per i coscritti non selezionati di Italia, Francia, Austria, Svizzera.

In Italia, gli iscritti alle leve di mare (l'iscrizione dipende dalla professione esercitata) sono sottoposti ad una doppia selezione, prima presso i Consigli di leva, poi presso i Depositi dei Corpi Equipaggi Marittimi.

Le stature estreme risultano già eliminate presso i Consigli di leva, cosicchè la distribuzione dei coscritti esaminati ai Depositi risulta iponormale. Per gli iscritti alle classi di leva 1876, 1879, 1882, esaminati al Deposito C.E.M. de La Spezia, l'indice risulta eguale a 3,068.

L'iponormalità si accentua solo lievemente tra i soldati dichiarati abili dai detti depositi ($I = 3,062$).

È interessante esaminare separatamente le distribuzioni degli abili nelle tre classi 1876, 1879, 1882, in quanto diversi furono i criteri di reclutamento per esse adottate, meno severi per la classe 1879 che per le classi 1876 e 1882. Se ne avverte l'effetto sull'indice che accusa una iponormalità alquanto minore per

(1) Volume citato nella nota (15), pagg. 8, 12, 14, 17, 20, 23, 48.

quella che per queste: ($I = 3,028$ per la classe 1876 ; 3,091 per la 1879 ; $= 3,027$ per la 1882).

Dal Quetelet sono state pubblicate anche due serie di dati relative ai perimetri toracici di soldati scozzesi e americani (1). Per queste due distribuzioni, l'indice risulta, contrariamente a quanto avveniva per la statura, ipernormale, e precisamente uguale a 3,157 per gli Scozzesi e a 3,322 per gli Americani. La distribuzione dei perimetri toracici risulta in questo esempio, più ipernormale di quella delle stature.

Lo stesso fenomeno si verifica per le distribuzioni relative alla statura e al perimetro toracico di 2600 reclute finniche negli anni 1901, 1902 e 1903 (2). L'indice, per la distribuzione delle stature, è 3,126 ; per la distribuzione dei perimetri toracici è 3,310.

È stato anche calcolato l'indice d'ipernormalità della distribuzione relativa al peso di questi soldati e si è trovato che questa distribuzione è più ipernormale della distribuzione relativa alla statura e meno ipernormale della distribuzione relativa al perimetro toracico, l'indice risultando uguale a 3,216.

Nella tav. IX-*bis* gli indici di ipernormalità per i perimetri toracici e i pesi dei militari in servizio dei detti paesi sono posti a confronto con quelli già sopra ricordati per le stature.

La conclusione che se ne trae è confermata dal calcolo degli indici di ipernormalità (dati nella tav. IX-*ter*) per le distribuzioni delle 49.054 reclute belghe dichiarate abili (vedi tavola VIII-*bis*). Mentre per le stature l'indice presenta, come abbiamo visto, un valore ($= 3,138$) lievemente inferiore al normale, per il perimetro toracico e per il peso i suoi valori sono nettamente ipernormali (e rispettivamente $= 3,244$ e $= 3,276$).

La spiegazione delle differenze tra gli indici ottenuti per gli uni e per le altre si può dare ammettendo che la maggiore ipernormalità di una distribuzione relativa ad un carattere antropometrico sia una proprietà intrinseca di tali caratteri. Ma si

(1) Per i 5738 soldati scozzesi, vedi *Fisica Sociale*, op. cit., pag. 651 ; per i 1516 soldati americani dell'armata del Potomac, Ibidem, pag. 652.

(2) KARL NICKUL. *Einige Körpermasse finnischer Wehrpflichtigen in mathematisch-statistischer Behandlung*, in « *Annales Academiae Scientiarum Fennicae* », Ser. A, Tom. XXV, Helsinki, 1927, pagg. 11, 12, 13.

potrebbe anche pensare che la selezione medica eserciti una diversa influenza sui vari caratteri.

Avendo però calcolato gli indici d'ipernormalità per le distribuzioni relative al peso e al perimetro toracico all'atto della visita medica e quindi prima della selezione, dei 71.664 belgi di cui già abbiamo calcolato l'indice per la statura, si sono trovati dei risultati che concordano con quelli precedenti. Infatti gli indici risultano (vedi tav. IX-ter) uguali, per la statura, a 3,165; per il peso, a 3,356 e, per il perimetro toracico, a 3,412. Pare dunque di dover accettare la prima spiegazione, pur non potendo escludere che anche la seconda abbia un certo fondamento.

I diagrammi alle figure 7, 7-bis, 7-ter mostrano la frequenza delle eliminazioni operate dalla visita medica secondo i valori della statura, del peso e del perimetro toracico (per le cifre, vedi la tav. VIII-bis). Essi spiegano come l'abbassamento dell'indice di ipernormalità in dipendenza della selezione sia minore per la statura, intermedio per il peso e maggiore per il perimetro toracico (vedi tav. VIII-ter).

Disponiamo anche di dati sulla distribuzione dei pesi e dei perimetri toracici per l'Italia, relativi sia agli iscritti delle leve di mare visitati al Deposito de La Spezia, sia sui volontari di 15-19 anni presentatisi per l'assunzione.

Sia per l'insieme delle quattro classi (15-19 anni), sia per le tre di 16-19 anni (escludendo la prima per la ragione esposta parlando delle stature), l'indice di ipernormalità risulta, così per il perimetro toracico come per il peso, superiore a quello per la statura (vedi tavola IX-quarter).

TAVOLA IX-quarter

*Indici di ipernormalità per i vari caratteri (statura, perimetro toracico, peso) dei volontari presentatisi per l'arruolamento della marina
(Dati degli « Annali di medicina navale »)*

ANNI DI ETÀ	STATURE	PERIMETRO TORACICO	PESI
15-19.	3.226	3.279	3.336
16-19.	3.228	3.344	3.395

Nella tav. IX-*quinquies* sono infine raccolti gli indici per gli iscritti alle leve di mare delle tre classi di leva del 1876, 1879 e 1882 visitati al Deposito C.E.M. de La Spezia dopo aver subito la selezione della visita presso i Consigli di leva. Gli indici, sia per il perimetro toracico, sia e, più ancora, per il peso, risultano notevolmente più elevati che per la statura. Lieve influenza sugli indici

TAVOLA IX *quinquies*

Indici di ipernormalità per vari caratteri (statura, perimetro toracico, peso) degli iscritti alle leve di mare (classi 1876-1879-1882) visitati al Deposito corpi equipaggi marittimi de La Spezia.

	STATURE	PERIMETRI TORACICI	PESI
Abili + Inabili (classi di leva 1876-1879-1882).	3.068	3.119	3.218
Abili (classi di leva 1876-1879-1882). . . .	3.062	3.105	3.217
Classi di leva			
Abili 1876.	3.028	3.043	3.225
Abili 1879.	3.091	3.121	3.261
Abili 1882.	3.027	3.089	3.172

di ipernormalità per questi caratteri, al pari che per la statura, esercita la successiva selezione intervenuta al Deposito C.E.M.

Per la statura, come anche per il peso e ancor più per il perimetro toracico, sensibile influenza sull'indice di ipernormalità esercitano invece le differenze nei criteri di reclutamento adottati per le diverse classi.

Contrariamente a quanto si è riscontrato per i militari finnici, belgi e americani (tav. IX-*bis* e IX-*ter*), la selezione riduce, per queste serie, a iponormale l'indice del perimetro toracico; restando invece ipernormale l'indice per il peso. Ciò si può spiegare con il fatto che, nel reclutamento per la marina, si accorda particolare importanza alle caratteristiche del perimetro toracico.

Per tutte le serie, ad ogni modo, è confermato che l'indice di ipernormalità risulta per la statura — carattere lineare —

nettamente inferiore che per il peso e il perimetro toracico, caratteri dipendenti da due o più dimensioni.

Per gli Stati Uniti d'America, furono pubblicate le distribuzioni delle stature per gli studenti (23.122 maschi e 17.127 femmine) di vari collegi americani ⁽¹⁾ in età da 16 a 21 anni.

L'elaborazione dei dati ha portato ai seguenti indici di ipernormalità :

Anni di età	MASCHI	FEMMINE
16	3,149	3,148
17	3,242	3,170
18	3,223	3,161
19	3,265	3,218
20	3,222	3,223
21	3,279	3,123
Media dei 6 indici.	3,233	3,174
Complessivamente per gli anni 6-21. . .	3,238	3,177

La media degli indici per gli studenti dei 6 anni non differisce sensibilmente dagli indici per il complesso degli studenti e tutti (salvo quelli delle studentesse di 21 anni) presentano ipernormalità. L'ipernormalità per tutti gli anni è meno accentuata per le donne che per gli uomini, salvo per i 20 anni per cui risulta praticamente uguale per i due sessi. È da ritenersi che una certa selezione si operi nel reclutamento degli studenti, gli individui eccezionali per alta o bassa statura rimanendo esclusi dall'università ed è verosimile che tale selezione sia più intensa per il sesso femminile, circostanza che può spiegare i più bassi indici di ipernormalità.

B. — *Dati relativi a popolazioni adulte.*

3. Sono state studiate le distribuzioni delle stature di alcune popolazioni adulte e in questo paragrafo riportiamo i relativi indici di ipernormalità.

⁽¹⁾ I collegi americani comprendono i due primi anni dell'università, corrispondenti ai due ultimi anni dei licei europei. I dati furono rilevati durante gli anni accademici 1928-29 e 1929-30 e si riferiscono a 10 Istituti per i maschi e a 8 per le femmine.

Vedi: HAROLD S., DIEHL, *Heights and Weights of American College Men*: «Human Biology», vol. V, n. 3, settembre 1933, pagg. 443-479; *Heights and Wights of American College Women*, «Human Biology», Vol. V, n. 4, dicembre 1933, pagg. 620-628.

Dallo scritto di S. Wahlund, *Anthropometry of the Swedish Lapps* ⁽¹⁾ sono stati estratti e quindi elaborati dati sulle distribuzioni delle stature dei Lapponi di età superiore ai 15 anni, distinti secondo il sesso. L'indice d'ipernormalità per la distribuzione relativa ai Lapponi maschi è = 3,374, mentre quello relativo alle donne lapponi è = 3,138. Sono stati poi divisi i Lapponi in tre gruppi di età e si sono ottenuti gli indici riportati nella tavola X.

Si nota che, in due casi su tre, le distribuzioni parziali hanno un indice d'ipernormalità maggiore di quello della distribuzione complessiva, sia per i maschi che per le femmine. Per le popolazioni relative a gruppi di età singole, si hanno quattro indici ipernormali e due iponormali, mentre per le popolazioni adulte complessive un indice è ipernormale e uno iponormale.

TAVOLA X

Indici di ipernormalità per la statura dei Lapponi adulti.

ETÀ	INDICI DI IPERNORMALITÀ	
	Uomini	Donne
15-31.	3.062	3.226
32-48.	3.485	2.994
49-85.	3.932	3.364
15-85.	3.374	3.138
Media dei primi tre indici	3.493	3.195

Possiamo quindi notare che le distribuzioni relative a gruppi di età singole risultano più spesso ipernormali.

Altri dati per popolazioni adulte complessive riguardano 39.465 uomini turchi e 20.263 donne turche e sono ricavati da un'apposita inchiesta eseguita da personale specializzato nel

(1) IN G. DAHLBERG and S. WAHLUND, *The Race Biology of the Swedish Lapps*. The Swedish State Institute of Human Genetics and Race Biology. Uppsala. Pagg. 49-81.

1937 ⁽¹⁾. Sono stati calcolati gli indici di ipernormalità per le due distribuzioni e si sono ottenuti i seguenti valori.

$$\text{UOMINI} \quad I = 3,226$$

$$\text{DONNE} \quad I = 3,170$$

I due indici superano quindi per entrambi i sessi, ma di poco per le donne, il valore normale.

Altri dati relativi alla statura di popolazioni adulte complessive sono stati ricavati da ricerche fatte dal professor Schlagin-

TAVOLA XI

Classificazione degli indici di ipernormalità per le stature di adulti di vari Paesi (Svizzera, Germania, Nuova Guinea, Micronesia, Isole Faeröer).

CLASSI	FREQUENZE		
	Maschi	Femmine	Totale
2,801 -2,900	3	1	4
2,901 -3,000	2	2	4
3,001 -3,100	2	3	5
3,101 -3,1416	4	2	6
3,1416-3,200	6	1	7
3,201 -3,300	0	1	1
3,301 -3,400	1	1	2
3,401 -3,500	0	1	1
TOTALE.	18	12	30
Media degli indici	3,084	3,116	3,097

(¹) I dati videro la luce nella pubblicazione n. 15 dell'Ufficio Centrale di statistica della Repubblica turca (*Enquête anthropométrique turque, faite sur 59.728 individus des deux sexes*, Ankara) e nella tesi di laurea in sociologia conseguita dalla Sig.na AFET UZMAY nell'Università di Ginevra (*Recherches sur les caractères anthropologiques des populations de la Turquie*, Faculté des Sciences économiques et sociales, Imprimerie Albert Künding, 1939).

haufen e da suoi discepoli su maschi e femmine di alcuni villaggi della Svizzera, della Germania e di isole della Melanesia e della Micronesia ⁽¹⁾. Dati sulle stature di adulti delle isole Faeröer sono stati estratti da una pubblicazione del Prof. F. Jørgensen ⁽²⁾.

Per i maschi, la media di 18 indici risulta uguale a 3,084, per le femmine la media di 12 indici è invece = 3,116.

Nella tavola XII riportiamo i valori degli indici per i vari paesi. L'indice che si riferisce a Bregenzer Wald è la media di 7

TAVOLA XII

Indici di ipernormalità per le stature di adulti di vari Paesi

P A E S I	INDICI	
	U o m i n i	D o n n e
ENGSTLIGEN UND FRUTIGTAL (Oberland, Svizzera)	3.162	3.172
UNTERTOGGENBURG (S. Gallo, Svizzera)	3.097	3.093
SCHANGNAU (Oberemmenthal, Svizzera)	2.995	3.230
GENKINGEN (Germania)	3.122	3.122
BREGENZER-WALD (Germania)	3.098	3.063
NUOVA GUINEA (Melanesia)	3.069	—
LIR (Melanesia)	3.19	—
KAPINGAMARANGI (Micronesia)	2.85	—
FAERÖER (Danimarca)	3.18	3.33

indici relativi a distribuzioni delle stature di abitanti adulti di Bregenzer Wald e di paesi vicini. L'indice che si riferisce alla Nuova Guinea è la media di 4 indici relativi a quattro villaggi della costa settentrionale (Torricelligebirge, Yakumul, Arup, Leitere). Si può osservare che, contrariamente a quanto avveniva per le distribuzioni relative ai coscritti, c'è una prevalenza degli

⁽¹⁾ Nel prossimo capitolo verranno specificati i titoli delle pubblicazioni da cui sono tratti questi dati e quelli relativi ad altri caratteri antropometrici di queste popolazioni (Vedi pagg. 39-44).

⁽²⁾ Op. cit. alla nota 29 a pag. 45.

indici iponormali su quelli ipernormali e precisamente, per i maschi, 11 indici sono iponormali e 7 ipernormali e, per le femmine, 8 indici sono iponormali e 4 ipernormali.

Sono state esaminate separatamente alcune distribuzioni sulla statura di abitanti adulti della Nuova Irlanda ⁽¹⁾. Gli indici di

TAVOLA XIII

Classificazione degli indici di ipernormalità per le stature della Nuova Irlanda.

CLASSI	FREQUENZE		
	Maschi	Femmine	Totale
-2,600	1	—	1
2,600 -2,700	1	1	2
2,701 -2,800	1	—	1
2,801 -2,900	2	1	3
2,901 -3,000	2	1	3
3,001 -3,100	4	2	6
3,101 -3,1416	1	—	1
3,1417-3,200	6	1	7
3,201 -3,300	5	1	6
3,301 -3,400	3	2	5
3,401 -3,500	2	—	2
3,501 -3,600	1	1	2
3,601 -3,700	1	—	1
3,701 -3,800	—	1	1
oltre 3,800	1	—	1
TOTALE	31	11	42
MEDIA DEGLI INDICI	3,180	3 176	—

(1) OTTO SCHLAGINHAUFEN, *Die Variabilität, geographische Verteilung und Stellung der Körpergrösse der Eingeborenen Neuirlands*. «Geographica Helvetica», Heft 1, 1953.

ipernormalità di queste distribuzioni risultano in prevalenza ipernormali, (19 ipernormali contro 12 iponormali per i maschi e 6 contro 5 per le femmine), con una media di 3,180 per i maschi e 3,176 per le femmine. Riportiamo nella tavola XIII la classificazione degli indici in parola.

DATI PER ALTRI CARATTERI.

4. La ricerca che si era effettuata sulle distribuzioni delle stature e, in alcuni casi, dei pesi e dei perimetri toracici, sia di coscritti, che di popolazioni adulte, venne estesa a molteplici altri caratteri antropometrici.

Particolarmente utili sono state, per questo studio, le indagini del prof. Schlaginhaufen e dei suoi allievi su molteplici caratteri antropometrici di coscritti svizzeri ⁽¹⁾ e di abitanti adulti di villaggi appartenenti a vari paesi (Svizzera, Germania, isole della Melanesia e della Micronesia).

Gli indici di ipernormalità per le distribuzioni relative a 34 caratteri antropometrici dei coscritti svizzeri per il complesso

TAVOLA XIV

Classificazione degli indici di ipernormalità per 34 caratteri relativi agli iscritti di leva per il complesso della Confederazione Svizzera.

CLASSI	FREQUENZE
3,11 -3,1416	—
3,1416-3,20	5
3,21 -3,30	19
3,31 -3,40	8
3,41 -3,50	1
3,51 -3,60	1
TOTALE	34
INDICE MEDIO DI IPERNORMALITÀ	3.275

⁽¹⁾ Vedi pagg. 10-11.

della Confederazione hanno dato tutti valori superiori a 3,1416 con una media di 3,275.

Per i singoli cantoni svizzeri, invece, l'indice risulta a volte, forse in dipendenza dello scarso numero di osservazioni, inferiore al valore normale, ma questo succede solo per 31 casi su 143.

La media dei 143 indici risulta 3,236, sostanzialmente concordante col valore medio 3,275 trovato per la distribuzione relativa alla Confederazione. Resta così confermato, anche per gli altri caratteri antropometrici, ciò che era risultato per le stature, e precisamente che, frazionando la circoscrizione territoriale, non si ottiene una riduzione sostanziale della ipernormalità.

TAVOLA XV

Classificazione degli indici di ipernormalità relativi agli iscritti di leva per i singoli Cantoni Svizzeri.

CLASSI	FREQUENZE
2,801 -2,900	1
2,901 -3,000	—
3,001 -2,100	14
3,101 -3,1416	16
3,1411-3,200	25
3,201 -3,300	47
3,301 -3,400	31
3,401 -3,500	6
3,501 -3,600	1
3,601 -3,700	1
3,701 -3,800	—
3,801 -3,900	1
3,901 -4,000	—
TOTALE	143
INDICE MEDIO DI IPERNORMALITÀ	3,236

Da quello che abbiamo visto sopra deriva, che, anche per altri molteplici caratteri antropometrici dei coscritti, si verifica come per le stature, che gli indici di ipernormalità superano 3,1416. Se invece per gli stessi caratteri esaminiamo distribuzioni relative a popolazioni adulte senza distinzione di età, molto spesso troviamo indici inferiori a 3,1416, come passiamo a mostrare.

TAVOLA XVI

Media degli indici di ipernormalità per alcuni caratteri antropometrici di coscritti svizzeri e delle popolazioni adulte maschili e femminili

CARATTERE	MEDIA DEGLI INDICI DI IPERNORMALITÀ		
	Coscritti svizzeri	Popolazione adulta maschile di villaggi della Svizzera e di altri paesi	Popolazione adulta femminile di villaggi della Svizzera e di altri paesi.
Statura	3.278 (25)	3.084 (15)	3.116 (12)
Lunghezza del tronco relat.	3.359 (3)	3.121 (9)	3.231 (8)
Apertura Braccia relat.	3.373 (3)	3.140 (18)	3.146 (12)
Lungh. massima testa	3.195 (8)	3.122 (16)	3.110 (10)
Lungh. capo relativa.	3.270 (5)	3.121 (10)	3.089 (8)
Largh. massima testa	3.203 (8)	3.061 (16)	3.149 (11)
Indice cefalico.	3.214 (14)	3.113 (17)	3.100 (10)
Altezza Morfol. Faccia	3.221 (3)	3.152 (14)	3.127 (10)
Diametro bizigomatico	3.190 (7)	3.196 (16)	3.017 (10)
Diametro bigoniaco	3.230 (7)	3.196 (9)	3.128 (9)
Indice morf. faccia.	3.208 (8)	3.253 (12)	3.061 (10)
Altezza del naso.	3.190 (7)	3.121 (18)	3.105 (12)
Larghezza del naso	3.172 (7)	3.232 (18)	3.126 (12)
Indice nasale	3.219 (8)	3.124 (17)	3.161 (11)
Indice jugo-mandibolare	3.149 (6)	3.237 (11)	3.224 (11)
Diametro frontale minimo	3.385 (1)	3.139 (1)	3.080 (1)
Altezza fisionomica della faccia super. {	Diam. ophryo-alveolaris Altezza nasion-stomion	3.580 (1)	3.131 (1)
		3.308 (4)	

Lo Schlaginhaufen e i suoi allievi, oltre ad eseguire misure di caratteri antropometrici per coscritti svizzeri hanno infatti eseguito, come abbiamo già detto, indagini sopra gli adulti di villaggi appartenenti a vari paesi tra cui anche alla Svizzera.

Ora, se, per i singoli caratteri, si confrontano le medie degli indici di ipernormalità ottenuti per i coscritti svizzeri e per gli adulti dei detti villaggi, si trova che i valori per questi ultimi sono quasi sempre inferiori. Nella tavola XVI sono illustrati questi confronti. Le cifre tra parentesi indicano il numero degli indici di ipernormalità di cui si è fatta la media.

I confronti si riferiscono a 17 caratteri : per essi, gli indici medii di ipernormalità ottenuti per i coscritti svizzeri si possono contrapporre agli indici medii di ipernormalità risultanti per le popolazioni adulte femminili e per le maschili dei villaggi considerati. Si

TAVOLA XVII

Classificazioni degli indici di ipernormalità per la popolazione adulta di Engstligen und Frutigtal (Oberland, Svizzera).

CLASSI	FREQUENZE		
	Maschi	Femmine	Totale
<2,801	3	—	3
2,801 -2,900	2	2	4
2,901 -3,000	5	4	9
3,001 -3,100	9	14	23
3,101 -3,1416	6	8	14
3,1416-3,200	5	1	6
3,201 -3,300	2	6	8
3,301 -3,400	2	4	6
3,401 -3,500	3	2	5
3,501 -3,600	3	—	3
>3,600	1	—	1
TOTALE	41	41	82
MEDIA DEGLI INDICI	3.106	3.136	3.121

hanno così 34 paragoni, in 28 dei quali il valore ottenuto per i coscritti svizzeri risulta superiore a quello ottenuto per le popolazioni adulte dei villaggi.

Sofferamiamoci sui risultati ottenuti per i villaggi svizzeri. Essi sono nettamente diversi da quelli ottenuti per i coscritti.

Dei tre villaggi della Svizzera, in quello di Engstligen e Frutigtal nell'Oberland ⁽¹⁾ su 41 indici, 25 per i maschi, 27 per le femmine sono minori di 3,1416; le medie risultano = 3,106 per i maschi e = 3,136 per le femmine.

Nella tavola XVII sono classificati gli indici di ipernormalità per le distribuzioni, sia dei maschi che delle femmine di detto villaggio.

TAVOLA XVIII

Classificazione degli indici di ipernormalità per le popolazioni adulte di Untertoggenburg (S. Gallo, Svizzera).

CLASSI	FREQUENZE		
	Maschi	Femmine	Totale
2,501 -3,000	4	10	14
3,010 -3,100	12	7	19
3,101 -3,1416	3	1	4
3,1416-3,200	4	4	8
3,201 -3,300	9	8	17
3,301 -3,400	5	6	11
3,401 -3,500	5	1	6
3,501 -3,600	2	6	8
3,601 -3,700	1	2	3
3,701 -3,800	1	1	2
TOTALE.	46	46	92
MEDIA DEGLI INDICI.	3.221	3.212	3.216

⁽¹⁾ O. SCHLAGINHAUFEN, *Anthropologische Untersuchungen im Engstligen- und Frutigtal (Berner Oberland)*. Aus dem anthropologischen Institut der Universität Zürich. Orell Füssli, Zürich, 1938.

Nel circondario di Untertoggenburg (cantone di S. Gallo ⁽¹⁾), la frequenza è maggiore per gli indici ipernormali che per quelli iponormali; per i maschi, vi sono 19 indici iponormali e 27 ipernormali; per le femmine, 18 iponormali e 28 ipernormali. La media è 3,221 per i maschi e 3,212 per le femmine. Riportiamo la classificazione degli indici nella tavola XVIII.

Nel villaggio di Schangnau nell'Oberemmental ⁽²⁾, per gli uomini, gli indici iponormali sono 25 contro 18 ipernormali; per

TAVOLA XIX

Classificazione degli indici di ipernormalità per la popolazione adulta di Schangnau (Oberemmental, Svizzera).

CLASSI	FREQUENZE		
	Maschi	Femmine	Totale
2,500 -3,000	13	12	25
3,001 -3,100	7	7	14
3,101 -3,1416	5	2	7
3,1416-3,200	2	6	8
3,201 -3,300	2	7	9
3,301 -3,400	3	6	9
3,401 -3,500	4	—	4
3,501 -3,600	1	—	1
3,601 -3,700	1	—	1
3,701 -3,800	2	—	2
3,801 3,900	—	1	1
oltre 3,900	3	2	5
TOTALE.	43	43	86
MEDIA DEGLI INDICI	3.210	3.177	3.193

⁽¹⁾ O. SCHLAGINHAUFEN, *Anthropologische Untersuchungen im Bezirk Untertoggenburg (Kt. St. Gallen, Schweiz)*. Aus dem anthropologischen Institut der Universität Zürich.

⁽²⁾ O. SCHLAGINHAUFEN, *Zur Kenntnis der Bevölkerung von Schangnau im Oberemmental*. « Mitteilungen der Geographisch-Ethnographischen Gesellschaft », Band XXXIX (1938-39), Zürich, 1939.

le donne, 23 contro 20 ipernormali. La media degli indici è 3,210 per i maschi e 3,177 per le femmine. Riportiamo la classificazione degli indici nella tavola XIX.

Qui, come in altri casi che vedremo in seguito, le medie degli indici risultano ipernormali anche per i villaggi in cui i singoli indici sono più spesso iponormali. Ciò avviene perchè la distribuzione dei singoli indici non è simmetrica, ma gli indici ipernormali eccedono sul valore normale più di quanto ne restino al di sotto gli indici iponormali.

Nei villaggi della Germania, si presenta invece una fortissima prevalenza degli indici iponormali: nel villaggio di Genkingen (Jena) ⁽¹⁾ vi sono 21 indici iponormali, contro uno ipernormale per i maschi e 19 contro 3 per le femmine, con medie rispettivamente di 3,127 e 3,147. A Bregenzer Wald e località

TAVOLA XX

Classificazione degli indici di ipernormalità per la popolazione adulta di Genkingen (Germania).

CLASSI	FREQUENZE		
	Maschi	Femmine	Totale
3,101 -3,1416	21	19	40
3,1416-3,200	1	2	3
3,201 -3,300	—	—	—
3,301 -3,400	—	—	—
3,401 -3,500	—	—	—
3,501 -3,600	—	1	1
3,601 -3,700	—	—	—
TOTALE.	22	22	44
MEDIA DEGLI INDICI	3.127	3.147	3.137

⁽¹⁾ A. BREIG, *Eine anthropologische Untersuchung auf der Schwäbischen Alb (Dorf Genkingen)*, « Deutsche Rassenkunde », Jena G. Fischer, 1935.

vicine ⁽¹⁾ gli indici iponormali sono 93 contro 43 ipernormali per i maschi e 102 contro 34 per le femmine, con medie rispettivamente di 3,118 e 3,117. Riportiamo le classificazioni degli indici per i due villaggi della Germania nelle tavole XX e XXI.

TAVOLA XXI

Classificazione degli indici di ipernormalità per la popolazione adulta di Bregenzer Wald e località vicine (Germania).

CLASSI	FREQUENZE		
	Maschi	Femmine	Totale
2,401 -2,500	1	—	1
2,501 -2,600	1	1	2
2,601 -2,700	2	1	3
2,701 -2,800	4	1	5
2,801 -2,900	8	10	18
2,901 -3,000	13	10	23
3,001 -3,100	13	20	33
3,101 -3,1416	51	59	110
3,1416-3,200	19	13	32
3,201 -3,300	9	14	23
3,301 -3,400	7	1	8
3,401 -3,500	3	2	5
3,501 -3,600	2	2	4
3,601 -3,700	—	1	1
3,701 -3,800	—	1	1
3,801 -3,900	1	—	1
oltre 3,900	2	—	2
TOTALE	136	136	272
MEDIA DEGLI INDICI	3.118	3.117	3.117

⁽¹⁾ BRUNO K. SCHULTZ, *Bauern im südlichen Allgäu, Lechtal und Bregenzer Wald*, « Rassenkunde Deutscher Gaue » München, J. F. Lehmanns, 1935.

TAVOLA XXII

*Classificazione degli indici di ipernormalità per la popolazione
di 4 villaggi della Nuova Guinea.*

CLASSI	FREQUENZE
2.201 -2.300	1
2.301 -2.400	—
2.401 -2.500	—
2.501 -2.600	1
2.601 -2.700	6
2.701 -2.800	10
2.801 -2.900	8
2.901 -3.000	22
3.001 -3.100	15
3.101 -3.1416	7
3.1416-3.200	8
3.201 -3.300	11
3.301 -3.400	10
3.401 -3.500	7
3.501 -3.600	7
3.601 -3.700	6
3.701 -3.800	1
3.801 -3.900	2
3.901 -4.000	1
4.001 -4.100	—
4.101 -4.200	1
4.201 -4.300	—
4.301 -4.400	1
4.401 -4.500	—
TOTALE	125
MEDIA DEGLI INDICI	3,172

Nella Nuova Guinea ⁽¹⁾ e nelle isole di Lir (Melanesia) e Kapingamarangi (Micronesia) ⁽²⁾, prevalgono gli indici iponormali (70 contro 55 ipernormali e, rispettivamente, 21 contro 18).

TAVOLA XXIII

Classificazione degli indici di ipernormalità per la popolazione di Lir (Melanesia) e Kapingamarangi (Micronesia).

CLASSI	FREQUENZE
2,601 -2,700	1
2,701 -2,800	4
2,801 -2,900	1
2,901 -3,000	8
3,001 -3,100	4
3,101 -3,1416	3
3,1416-3,200	3
3,201 -3,300	2
3,301 -3,400	4
3,401 -3,500	4
3,501 -3,600	2
3,601 -3,700	—
3,701 -3,800	1
3,801 -3,900	1
oltre 3,900	1
TOTALE	39
MEDIA DEGLI INDICI	3,143

⁽¹⁾ O. SCHLAGINHAUFEN, *Mittelwert (M), stetige Abweichung (σ) und Variationskoeffizient (c) der an den vier Neuguineagruppen beobachteten Merkmale, Anthropometrische Untersuchungen an Eingeborenen in Deutsch-Neuguinea*, Leipzig, G. G. Teubner, 1914.

⁽²⁾ O. SCHLAGINHAUFEN, *Zur Anthropologie der mikronesischen Inselgruppe Kapingamarangi (Greenwich-Inseln)*, « Archiv der Julius Klaus-Stiftung », Band IV, Haft 3, Zurich, Orell Füssli, 1929.

La media degli indici è 3,143 nella Nuova Guinea e 3,172 nelle due isole. Riportiamo nelle tavole XXII e XXIII la classificazione degli indici.

Da tutto ciò che abbiamo visto, appare chiaro che in generale le distribuzioni relative a tutte le età adulte risultano, se non iponormali, quanto meno in minimo grado ipernormali in confronto alle distribuzioni relative ai coetanei. Entra dunque in azione un fattore che tende ad abbassare o annullare la ipernormalità, fattore che agisce però in maniera diversa sui diversi caratteri e nei diversi paesi.

Quest'ultima conclusione ci viene confermata dalla elaborazione dei dati ottenuti da altri autori per altre popolazioni: Eschimesi ⁽¹⁾, abitanti delle isole Faeröer ⁽²⁾, Samaritani ⁽³⁾ (per

TAVOLA XXIV

Classificazione degli indici di ipernormalità per gli Eschimesi.

CLASSI	FREQUENZE
2.800 -2.900	1
2.901 -3.000	5
3.001 -3.100	5
3.101 -3.1416	0
3.1416 -3.200	1
3.201 -3.300	1
3.301 -3.400	2
3.401 -3.800	2
TOTALE.	17
MEDIA DEGLI INDICI.	3.105

(1) KAJ BIRKET SMITH, *Report of the fifth Thule Expedition 1921-24*, Vol. III, No. 2, *Anthropological observations on the Central Eskimos*, Copenhagen, 1940.

(2) F. JÖRGENSEN, *Anthropologiske Undersøgelser fra Faerøerne (Anthropologia faeroica)*.

(3) GIUSEPPE GENNA, *I Samaritani*, Comitato Italiano per lo studio dei problemi della popolazione. Spedizioni scientifiche dirette dal Prof. C. Gini, Roma, 1938.

TAVOLA XXV

Classificazione degli indici di ipernormalità per i crani di maschi adulti di Teneriffa.

CLASSI	FREQUENZE			
	Giovani	Età media	Anziani	Popolazione complessiva adulta
2,701 -2,800	1	—	4	—
2,801 -2,900	3	2	1	1
2,901 -3,000	3	1	4	3
3,001 -3,100	7	7	6	6
3,101 -3,1416	1	3	5	4
1,416 -3,200	3	4	1	4
3,201 -3,300	5	5	1	5
3,301 -3,400	1	4	2	3
3,401 -3,500	3	1	1	1
3,501 -3,600	1	1	2	1
3,601 -3,700	—	1	—	—
3,701 -3,800	—	—	—	—
oltre 3,800	—	—	2	1
TOTALE	29	29	29	29
MEDIA DEGLI INDICI	3.193	3,201	3.103	3.205

molteplici caratteri dei viventi), antichi abitanti delle isole Canarie ⁽¹⁾ (per i caratteri dei crani).

Per le distribuzioni relative ai caratteri antropometrici degli abitanti delle isole Faeröer e per quelle relative ai caratteri dei crani degli abitanti, maschi e femmine, di Teneriffa, si ha una uguale frequenza degli indici iponormali e ipernormali; per le distribuzioni relative ai caratteri dei crani di Gomera e del complesso delle altre Isole Canarie e per le distribuzioni relative ai caratteri antropometrici degli Eschimesi, è maggiore la frequenza

⁽¹⁾ ERNEST A. HOOTON, *The Ancient Inhabitants of the Canary Islands* Harvard African Studies. Volume VII, Cambridge, Mass., 1925.

degli indici iponormali ; per le distribuzioni che si riferiscono alle donne e agli uomini samaritani, è invece maggiore la frequenza degli indici ipernormali.

Salvo per gli Eschimesi e per i maschi di Gomera e del complesso delle altre isole Canarie, la media degli indici risulta però sempre ipernormale. (Pure ipernormali risultano, come abbiamo visto a pagg. 28-29, gli indici per le stature dei Lapponi e dei Turchi).

Su 17 indici di ipernormalità, relativi ai caratteri antropologici degli Eschimesi, 11 sono inferiori al valore normale e 6 superiori. La media degli indici è 3,105.

Riportiamo nella tavola XXIV la classificazione degli indici.

Esaminiamo ora gli indici di ipernormalità relativi ai 29 caratteri antropometrici dei crani di antichi abitanti adulti delle isole Canarie.

TAVOLA XXVI

Classificazione degli indici di ipernormalità per i crani di femmine adulte di Teneriffa.

CLASSI	FREQUENZE
2.700 -2.800	1
2.801 -2.900	3
2.901 -3.000	3
3.001 -3.100	7
3.101 -3.1416	1
3.1416-3.200	5
3.201 -3.300	2
3.301 -3.400	2
3.401 -3.500	1
3.501 -3.600	3
3.601 -3.700	1
TOTALE	29
MEDIA DEGLI INDICI	3,152

Nella tavola XXV è riportata la classificazione degli indici per la popolazione maschile adulta di Teneriffa.

Da questa classificazione risulta che, su 29 indici, 14 sono iponormali e 15 ipernormali. La media degli indici è 3,205. La popolazione è stata divisa in tre gruppi di età. Più sotto ne commenteremo i risultati.

Nella tavola XXVI è riportata la classificazione degli indici per la popolazione femminile di Teneriffa. Da questa classificazione risulta che 15 indici sono iponormali e 14 ipernormali.

La media degli indici è 3,152.

Per la popolazione maschile di Gomera, si hanno 26 indici iponormali e 3 ipernormali.

La media degli indici è 2,916.

Nella tavola XXVII è riportata la classificazione degli indici in parola.

TAVOLA XXVII

Classificazione degli indici di ipernormalità per i crani di maschi adulti di Gomera.

CLASSI	FREQUENZE
2.500 -2.600	1
2.601 -2.700	3
2.701 -2.800	4
2.801 -2.900	5
2.901 -3.000	10
3.001 -3.100	2
3.101 -3.1416	1
3.1416-3.200	—
3.201 -3.300	—
3.301 -3.400	2
3.401 -3.500	1
3.501 -3.600	—
TOTALE	29
MEDIA DEGLI INDICI	2,916

TAVOLA XXVIII

Classificazione degli indici di ipernormalità per i crani della popolazione maschile adulta complessiva delle quattro isole di Gran Canaria-Fuerteventura-La Palma-Hierro.

CLASSI	FREQUENZE
2.301 -2.400	1
2.401 -2.500	1
2.501 -2.600	0
2.601 -2.700	1
2.701 -2.800	2
2.801 -2.900	4
2.901 -3.000	6
3.001 -3.100	1
3.101 -3.1416	0
3.1416-3.200	4
3.201 -3.300	0
3.301 -3.400	2
3.401 -3.500	1
3.501 -3.600	1
3.601 -3.700	0
3.701 -3.800	1
3.801 -3.900	1
3.901 -4.000	1
4.001 -4.100	0
oltre -4.100	2
TOTALE.	29
MEDIA DEGLI INDICI	3,141

Per i crani della popolazione maschile di Gran Canaria, Fuerteventura, La Palma e Hierro si hanno 16 indici iponormali e 15 ipernormali.

La media degli indici è 3,141.

Nella tavola XXVIII, riportiamo la classificazione degli indici.

Per le distribuzioni relative alla popolazione adulta maschile di tutte e sei le isole considerate, si hanno 14 indici iponormali e 15 ipernormali.

La media degli indici è 3,164.

Nella tavola XXIX riportiamo la classificazione degli indici.

TAVOLA XXIX

Classificazione degli indici di ipernormalità per i crani della popolazione adulta maschile complessiva delle sei isole Teneriffa-Gomera-Gran Canaria-Fuerteventura-La Palma-Hierro.

CLASSI	FREQUENZE
2.901 -3.000	2
3.001 -3.100	5
3.101 -3.1416	7
3.1416-3.200	5
3.201 -3.300	7
3.301 -3.400	2
3.401 -3.500	1
3.501 -3.600	—
TOTALE.	29
MEDIA DEGLI INDICI	3,164

I dati sui Samaritani sono stati ricavati dai risultati ottenuti dal professor Genna in una spedizione scientifica diretta dal professor Corrado Gini. ⁽¹⁾. Nella tavola XXX riportiamo la classifi-

⁽¹⁾ Vedi nota (3) pag. 45.

cazione degli indici di ipernormalità per la popolazione maschile e femminile. Per la popolazione maschile, si hanno 20 indici iponormali contro 35 ipernormali; per quella femminile, 13 indici iponormali contro 22 ipernormali; nel primo caso, la media degli indici è 3,278, nel secondo è 3,271.

TAVOLA XXX

Classificazione degli indici di ipernormalità per i Samaritani.

CLASSE	FREQUENZE		
	Maschi	Femmine	Totale
2.701 -2.800	2	0	2
2.801 -2.900	3	3	6
2.901 -3.000	3	5	8
3.001 -3.100	7	3	10
3.101 -3.1416	5	2	7
3.1416-3.200	6	3	9
3.201 -3.300	7	3	10
3.301 -3.400	6	5	11
3.401 -3.500	3	3	6
3.501 -3.600	0	1	11
3.601 -3.700	4	3	7
3.701 -3.800	8	3	11
3.801 -3900.	0	1	1
3.901 -4.000	1	1	2
TOTALE.	55	36	91
MEDIA DEGLI INDICI.	3.278	3.271	3.275

Esaminiamo ora gli indici relativi ai caratteri antropometrici degli abitanti delle isole Faeröer. Come si vede dalla classificazione degli indici riportata nella tavola XXXI, su 11 indici, 5 per i maschi e 6 per le femmine sono iponormali. La media degli indici, per i maschi, è 3,172; per le femmine; 3,150.

Classificazione degli indici di ipernormalità per la popolazione delle Isole Faeröer.

CLASSI	FREQUENZE		
	Maschi	Femmine	Totale
3,000 -3,100	3	5	8
3,101 -3,1416	2	1	3
3,1416-3,200	4	1	5
3,201 -3,300	1	2	3
3,301 -3,400	0	2	2
3,401 -3,500	0	0	0
3,501 -3,600	1	0	1
TOTALE.	11	11	22
MEDIA DEGLI INDICI	3.172	3.150	3.161

Molti dei dati relativi ai coscritti e ai soldati di cui abbiano esposto i risultati delle elaborazioni nelle pagine precedenti sono stati rilevati dai medici militari, in generale, però, (salvo forse alcuni relativi ai soldati della guerra di secessione americana) in tempi di pace.

I medici militari non hanno sempre, nelle loro rilevazioni, la meticolosa diligenza di uno scienziato, ma, trattandosi di misurazioni semplici, queste si possono accettare con confidenza quando, come di solito avviene in tempo di pace, i rilevatori non lavorano sotto la pressione dell'urgenza. Diverse sono le condizioni nei periodi di guerra e soprattutto allo scoppio della guerra e perciò le rilevazioni eseguite all'atto della coscrizione seguite ad una dichiarazione di guerra devono essere accettate con prudenza.

Mi limito a considerare i dati relativi ai reclutamenti dell'Inghilterra e degli Stati Uniti d'America nella prima guerra mondiale, i quali sono stati elaborati e pubblicati sotto la re-

sponsabilità di studiosi il cui nome costituisce di per sè una garanzia di serietà.

I dati sui 91.513 maschi da 20 a 21 anni esaminati dai « Medical Boards » della Gran Bretagna ispirano particolare fiducia, in quanto riguardano esami eseguiti in vista bensì della guerra, ma prima che questa scoppiasse. Sono stati elaborati dal Dr. W.J. Martin, dell'Ufficio di Ricerche statistiche del « Medical Research Council » (1).

Essi presentano un interesse tutto speciale perchè gli esaminati sono classificati a seconda del grado delle loro attitudini al servizio militare. Se ne distinguono quattro gradi. Nel primo, rientrano gli individui in condizioni normali di salute e di forza, capaci di sostenere l'esercizio fisico consentaneo alla loro età: in altre parole, gli individui completamente abili. Il secondo ed il terzo grado comprendono individui che presentano malattie o difetti, meno seri nel grado secondo, più gravi nel terzo, tali però che loro permettono di prestare qualche servizio. Nel quarto grado, infine, restano gli individui incapaci di rendere servizi di qualsiasi genere.

La tavola XXXI-bis porta, per ogni grado, oltre ai valori dell'indice di ipernormalità, anche quelli dello scostamento semplice medio 1S_A e dello scostamento quadratico medio 2S_A dalla media aritmetica, e ciò sia per la statura che per il peso e il perimetro toracico.

Resta confermata la più elevata ipernormalità delle distribuzioni del peso e del perimetro toracico in confronto a quelle della statura.

Resta pure confermato che, in confronto al complesso delle reclute, quelle dichiarate completamente abili (gruppo I) presentano indici di ipernormalità più bassi e quelle dichiarate completamente inabili (gruppo IV), invece, indici di ipernormalità più elevati.

È interessante rilevare come la variabilità, desunta sia dallo scostamento semplice medio che dallo scostamento quadratico medio, generalmente cresca con la gravità della limitazione fisica. Questo risultato può essere interpretato come una conferma della

(1) W. J. MARTIN, *The Physique of Young Adult Males*. Medical Research Council Memorandum No. 20, London, 1949.

TAVOLA XXXI-bis

GRADO	Numero dei casi	Statura (in pollici)			Peso (in libbre)			Perimetro toracico (in pollici)		
		¹ S _A	² S _A	I	¹ S _A	² S _A	I	¹ S _A	² S _A	I
1	2	3	4	5	6	7	8	9	10	11
I	74.429	2.02	2.58	3.25	12.46	16.10	3.34	1.52	1.94	3.26
II.	8.502	2.14	2.72	3.23	13.51	17.82	3.48	1.63	2.13	3.42
III	4.780	2.20	2.84	3.32	13.90	18.80	3.66	1.67	2.21	3.50
IV	2.183	2.47	3.21	3.36	14.61	19.34	3.50	1.74	2.27	3.40
Complessivamente.	91.084 ¹	2.05	2.63	3.27	12.76	16.57	3.37	1.55	2.00	3.33

¹ Compresi i casi non classificati per grado.

Indici di cograduazioni tra i dati alle

Statura	Peso	colonna Perimetro toracico
5 e 3 = 0,75	8 e 6 = 0,75	II e 9 = 0,25
5 e 4 = 0,75	8 e 7 = 0,75	II e 10 = 0,25

teoria del Quetelet e dei costituzionalisti italiani (De Giovanni, Viola) che nell'uomo medio vede anche il tipo dell'uomo sano.

Anche l'indice di ipernormalità, pur con qualche eccezione, mostra una tendenza a crescere con la gravità delle imperfezioni fisiche.

I dati pubblicati per l'America si fondano su rilevazioni di casi molto più numerosi, ma della cui attendibilità si ha ragione di dubitare, per quanto siano stati preparati, con la collaborazione del Maggiore A.G. Love, sotto la direzione del prof. Charles B. Davenport (¹), antropologo di larga fama. Senonchè evidentemente il Davenport non era responsabile della rilevazione dei dati.

Anche questi dati riguardano la statura, il peso e il perimetro toracico.

Essi si riferiscono a tutti i coscritti, ai riformati e agli smobilitati, ma le tre categorie contengono un numero molto diffe-

(¹) MEDICAL DEPARTMENT OF THE UNITED STATES ARMY IN THE WORLD WAR, Vol. XV, *Statistics*. Part one, *Army Anthropology* prepared under the direction of M.W. Ireland, Surgeon General of the Army by CHARLES B. DAVENPORT, Carnegie Institution of Washington, and ALBERT G. LOVE, Washington Government. Printing Office, 1921.

rente di casi. La prima riguarda il primo milione di reclute, la terza i due primi milioni, mentre la seconda si limita a un centinaio di migliaia di individui che, per il peso, scendono a poco più di 80.000.

Per gli smobilitati, si è presa anche la circonferenza dell'addome.

Scopo delle misure prese sugli smobilitati era di fornire indicazioni per il confezionamento delle uniformi, uno scopo che non esigeva necessariamente una grande precisione nelle misure.

Per la statura e il perimetro toracico, qualche volta la misura è indicata in centimetri, altre volte in pollici: per rendere comparabili gli indici di variabilità, le misure sono state ridotte tutte in centimetri.

Per il peso, si ha una classificazione più dettagliata relativa al complesso delle reclute e ai riformati in confronto a quella relativa agli smobilitati: per eliminare tale fattore di perturbazione, si è dovuto adottare per tutte e tre le categorie la classificazione più sommaria, ma si è potuto riscontrare che ciò, mentre non porta a differenze sensibili negli indici di variabilità, abbassa notevolmente gli indici di ipernormalità.

In tal modo i dati vengono resi comparabili formalmente; ma non pare che lo siano sostanzialmente. Tra gli smobilitati, la classe minima di peso è invero di 100-104 libbre, tra i coscritti (primo milione) di 95-99 libbre, tra i riformati relativi ai due primi milioni di coscritti di 89 libbre e meno. Mentre nella prima classe di coscritti, che pesano da 95 a 99 libbre, sono classificati 184 individui, tra i riformati nei primi due milioni di coscritti ve ne sarebbero ben 965 di tale peso, vale a dire oltre cinque volte il numero degli individui di tale peso nel primo milione di coscritti, ciò che è manifestamente impossibile.

Per gli smobilitati, la classificazione secondo il peso non comprende i riformati per piede piatto (che costituiscono una buona metà dei casi). Poichè i dati delle pubblicazioni lo permettevano, si sono eseguiti i calcoli per le stature e il perimetro toracico sia comprendendo che escludendo i riformati per piede piatto. I risultati non differiscono notevolmente.

La tavola XXXI-ter presenta i risultati ottenuti.

Dati i fattori di perturbazione sopra segnalati, non si può accordare ai confronti che un valore molto approssimativo.

TAVOLA XXXI-ter

CATEGORIE DI MILITARI	NUMERO DEI CASI	1S_A	2S_A	I
		STATURA (in cm.)		
Truppe smobilitate.	103.410	5.32	6.81	3.166
Totale coscritti	868.445	5.40	6.79	3.166
Riformati per difetti e malattie.	493.033	5.48	6.89	3.156 ⁴
PESO (in libbre)				
Truppe smobilitate.	81.558	13.16	16.52	3.308 ¹
Totale coscritti	868.445	13.92	17.54	3.170 ³
Riformati per difetti e malattie.	493.033	14.61	18.41	3.176 ^{5,6}
PERIMETRO TORACICO (in cm.)				
Truppe smobilitate.	101.918	4.12	5.17	3.156 ²
Totale coscritti	873.139	4.00	5.03	3.156
Riformati per difetti e malattie.	222.334	4.10	5.08	3.226

¹ 11 classi.

² Bianche: N. dei casi = 95.590; $^1S_A = 4.14$; $^2S_A = 5.18$; $I = 3.140$

Di colore: N. dei casi = 6.328; $^1S_A = 3.72$; $^2S_A = 4.87$; $I = 3.442$

³ 12 classi. Con la classificazione più dettagliata si ottengono i valori seguenti: $^1S_A = 13.64$; $^2S_A = 17.42$; $I = 3.262$.

⁴ Riformati per piede piatto; N. dei casi = 270.348; $^1S_A = 5.37$; $^2S_A = 6.75$; $I = 3.160$

Riformati per altri casi; N. dei casi = 222.685; $^1S_A = 5.61$; $^2S_A = 7.05$; $I = 3.160$

⁵ Riformati per piede piatto; N. dei casi = 270.348; $^1S_A = 13.24$; $^2S_A = 18.14$; $I = 3.242$

Riformati per altri casi; N. dei casi = 222.685; $^1S_A = 14.26$; $^2S_A = 18.15$; $I = 3.236$

⁶ 13 classi. Con la classificazione più dettagliata si ottengono i valori seguenti: $^1S_A = 14.33$;

$^2S_A = 18.30$; $I = 3.262$.

Risulta confermato per i tre caratteri (statura, peso e perimetro toracico), la maggiore variabilità dei riformati di fronte al complesso dei coscritti, al pari che la maggiore variabilità di questi rispetto agli smobilitati per ciò che concerne la statura ed il peso. Più variabili rispetto alle altre due categorie, risultano invece gli smobilitati per ciò che riguarda il perimetro toracico. Si potrebbe forse pensare che le fatiche di guerra abbiano reso i soldati più differenziati di quanto lo erano all'atto delle coscrizioni, sviluppando maggiormente il perimetro toracico degli individui robusti e facendo invece deperire i più deboli che mal sopportavano le fatiche di guerra.

In contrasto con quanto risultava per i dati inglesi, non si riscontrano, invece, regolari variazioni dall'una all'altra categoria per quanto riguarda gli indici di ipernormalità. Per il peso e il perimetro toracico, l'ipernormalità risulta infatti maggiore per i riformati che per il totale delle reclute, ma per la statura essa risulta invece minore. Gli smobilitati presentano per il peso, ipernormalità maggiore del totale dei coscritti, ma l'indice appare uguale per le stature e il perimetro toracico.

Resta difficile rendersi ragione di questo comportamento. E dal pari difficile è rendersi conto del fatto che, se l'ipernormalità risulta per la statura alquanto minore che per il peso, essa risulta invece più elevata che per il perimetro toracico nel totale dei coscritti e negli smobilitati, in contrasto con i risultati di tutte le altre rilevazioni. L'ipotesi più verosimile è che queste discordanze dipendano dalla incompleta comparabilità, se non anche dalla scarsa attendibilità, dei dati.

L'indice in ogni modo accusa, per tutte le distribuzioni una più o meno sensibile ipernormalità.

Se, dall'esame dei singoli caratteri antropometrici, vogliamo passare a quello delle loro parti, non manchiamo di dati.

Avremo occasione di esporre più avanti (vedi tavole XXXV e XXXVI) gli indici di ipernormalità per le singole falangi delle varie dita, nonchè per le dita, che risultano dalla loro somma, e per i rispettivi metacarpi o metatarsi e, infine, per i raggi che risultano dalla somma delle singole dita e dei loro metacarpi o metatarsi.

Le due tavole XXXII e XXXII-*bis* danno, per la mano, l'una, e, per il piede, l'altra, la classificazione di tali indici, distinti a seconda che si tratta di uomo o di donna e dell'arto destro o del sinistro.

Le misurazioni, desunte dagli scheletri, sono state eseguite dallo Pfitzner e pubblicate nel lavoro più innanzi citato (vedi nota 1 a pag. 86).

Sia nel complesso, sia nelle singole classificazioni, prevalgono i valori ipernormali salvo per la mano della donna. Viceversa, per le ossa del piede, i valori ipernormali sono più frequenti nella donna che nell'uomo. Non appare facile dare la spiegazione di queste differenze.

TAVOLA XXXII

Classificazione degli indici di ipernormalità delle ossa della mano

VALORI DEGLI INDICI DI IPERNORMALITÀ	Uomo			Donna			Uomo + Donna		
	D	S	D + S	D	S	D + S	D	S	D + S
2,00				3		3	3		3
2,50					1	1		1	1
2,75 - 2,80.				1	2	3	1	2	3
2,81 - 2,90.	1		1	3	3	6	4	3	7
2,91 - 3,00.	1	2	3	4	5	9	5	7	12
3,01 - 3,10.	4	2	6	5	6	11	9	8	17
3,11 - 3,1416.		4	4	2		2	2	4	6
3,1416 - 3,20.		7	7	1	2	3	1	9	10
3,21 - 3,30.	7	7	14	4	3	7	11	10	21
3,31 - 3,40.	8	3	11	1	2	3	9	5	14
3,41 - 3,50.	4	1	5	3	2	5	7	3	10
3,51 - 3,60.	2	1	3	1	2	3	3	3	6
3,61 - 3,70.		1	1					1	1
3,71 - 3,80.		1	1	1		1	1	1	2
3,81 - 3,90.					1	1		1	1
3,91 - 4,00.	2		2				2		2
4,01 - 4,10.									
- 3,1416.	6	8	14	18	17	35	24	25	49
3,1416 e più	23	21	44	11	12	23	34	33	67

TAVOLA XXXII-bis

Classificazione degli indici di ipernormalità delle ossa del piede

VALORI DEGLI INDICI DI IPERNORMALITÀ	UOMO			DONNA			UOMO + DONNA		
	D	S	D + S	D	S	D + S	D	S	S + D
2,00					1	1		1	1
2,52 - 2,53	2		2				2		2
2,75	1		1				1		1
2,81 - 2,90	2	2	4	2	1	3	4	3	7
2,91 - 3,00	1	5	6	4	3	7	5	8	13
3,01 - 3,10	2	4	6	1	3	4	3	7	10
3,11 - 3,1416	3	3	6	2	1	2	5	3	8
3,1416 - 3,20	9	3	12		1	1	9	4	13
3,21 - 3,30	3	5	8	3	3	6	6	8	14
3,31 - 3,40	4	3	7	4	2	6	8	5	13
3,41 - 3,50	1	2	3	5	3	8	6	5	11
3,51 - 3,60				2	6	8	2	6	8
3,61 - 3,70		2	2	2	2	4	2	4	6
3,71 - 3,80	1		1	1	1	2	2	1	3
3,81 - 3,90									
3,91 - 4,00					2	2		2	2
4,01 - 4,10									
4,11 - 4,20				1		1	1		1
4,21 - 4,30				1		1	1		1
4,31 - 4,40					1	1		1	1
4,71 - 4,80				1		1	1		1
- 3,1416	11	14	25	9	8	17	20	22	42
3,1416 e più	18	15	33	20	21	41	38	36	74

Anche nella tavola XXXVII, relativa alle famiglie numerose del Comune di Pisino (Istria), vi sono, oltre agli indici di ipernormalità per la statura e per l'apertura delle braccia, quelli per varie parti del corpo: oltre agli indici per la statura seduta e per l'arto inferiore, quelli per due componenti della statura seduta (lunghezza della faccia e lunghezza del tronco) e per le componenti della lunghezza del tronco (lunghezza sternale, lunghezza xifo-epigastrica e lunghezza epigastrico-pubica). I dati sono distinti a seconda del sesso: 7 su 9, per ciò che riguarda gli uomini, e 8 su 9, per ciò che riguarda le donne, risultano ipernormali. Per quelli, la media aritmetica risulta = 3,424 e, per queste, = 3,342.

Abbiamo detto sopra che esiste un fattore che tende ad abbassare la ipernormalità delle distribuzioni relative ai caratteri antropologici degli individui di tutte le età adulte rispetto a quelle relative a coetanei. Questo fattore potrebbe essere la variazione dei valori medi con l'età. Una distribuzione relativa ad individui di tutte le età adulte può, infatti, risultare dalla somma di più distribuzioni normali, aventi, in corrispondenza alle varie età, medie diverse e variabilità analoghe: ora una tale distribuzione è appunto iponormale ⁽¹⁾. Per verificare questa ipotesi, è stata operata la ripartizione in tre classi (giovani, di età media, di età avanzata) per alcune distribuzioni relative a caratteri antropometrici di popolazioni adulte e precisamente per le distribuzioni delle stature dei Lapponi e di alcuni caratteri antropometrici dei crani di antichi abitanti adulti di Teneriffa.

I dati relativi alla stature dei Lapponi sono stati già riportati nella tavola X e abbiamo notato a suo luogo che la frequenza degli indici ipernormali, in confronto a quelli iponormali è più spesso superiore che inferiore per le singole classi in confronto alla popolazione complessiva: precisamente quattro su sei sono gli indici ipernormali per le distribuzioni parziali e uno su due per quelle complessive, risultato che è conforme all'ipotesi suddetta.

Per quanto riguarda, invece, i crani di Teneriffa, non si nota che gli indici ipernormali siano più numerosi di quelli iponormali per le singole classi, in confronto alla popolazione complessiva

(1) Vedi, per un caso particolare, Appendice prima.

(Tav. XXV). e l'indice di ipernormalità per la distribuzione complessiva (3,205) risulta superiore a tutte e tre le medie (3,193, 3,201, 3,103) delle distribuzioni parziali. Ciò si può attribuire al fatto che le medie delle misure e degli indici del cranio non variano sensibilmente con l'età. È inoltre da tenere presente che la decomposizione è stata fatta in classi ampie, dato il numero limitato dei casi ⁽¹⁾.

Da questi risultati si può indurre che quel fattore, di cui abbiamo parlato, influisce in maniera diversa sui diversi caratteri, ciò che si spiega in quanto, per alcuni di essi, la variazione dei valori medi con l'età è insensibile; per altri, invece, è notevole.

INFLUENZA DELL'EREDITÀ E DELL'AMBIENTE SULLA IPERNORMALITÀ DELLE DISTRIBUZIONI DEI CARATTERI ANTROPOMETRICI.

5. Nel paragrafo 2 abbiamo confrontato le distribuzioni relative al peso, al perimetro toracico e alla statura di alcuni gruppi di coscritti e abbiamo notato che le distribuzioni del peso e del perimetro toracico erano più fortemente ipernormali delle distribuzioni della statura. Ci proponiamo ora di esaminare, se si può porre in relazione la ipernormalità della distribuzione relativa ad un carattere antropometrico con la influenza che hanno su di esso l'eredità e l'ambiente.

Da una pubblicazione del Prof. O. von Verschuer ⁽²⁾, abbiamo ricavato i dati esposti nella tavola XXXIII, che si riferiscono all'influenza dell'eredità e dell'ambiente sul peso, sul perimetro toracico, sulla statura e sulla lunghezza della testa.

La statura e la lunghezza della testa, come si vede chiaramente, dipendono moltissimo dall'eredità, mentre molto di meno risultano influenzati da essa il peso e il perimetro toracico. Ricordando che, per i coscritti e i militari, si è riscontrato che, per

(¹) Anche per le stature degli studenti, maschi e femmine, dai 16 ai 21 anni, dei Collegi degli Stati Uniti d'America, come per quelle degli iscritti alle leve di mare in Italia dai 15 ai 19 anni, non si osserva che gli indici di ipernormalità, per singoli anni di età, siano superiori all'indice per il complesso della età, e ciò per quanto la statura media varii con l'età. Ciò può dipendere dal fatto che la variabilità non è la stessa per gli individui dei vari gruppi di età, probabilmente anche in dipendenza dal numero ristretto di osservazioni.

(²) *Erbpathologie*, Zweite Auflage, Dresden, Steinkopf, 1937.

TAVOLA XXXIII

Influenza dell'eredità e dell'ambiente sui caratteri antropometrici.

CARATTERI ANTROPOMETRICI	A-INFLUENZA DELL'AMBIENTE	B-INFLUENZA DELL'EREDITÀ	A/B
Peso	85%	169%	1/2
Perimetro toracico	70%	165%	1/2,24
Statura	18,5%	192%	1/10,4
Lunghezza della testa	10%	72,5%	1/5,6

il peso e il perimetro toracico, l'indice di ipernormalità risulta superiore che per la statura (vedi pagg. 27-29), si è indotti a pensare che l'ambiente influenzi più dell'eredità l'ipernormalità della distribuzione e quindi che le distribuzioni dei caratteri che dipendono più dall'ambiente, siano più ipernormali.

Per controllare tale conclusione, vogliamo distinguere i caratteri antropometrici che abbiamo finora esaminati, a seconda che dipendono più o meno dall'eredità. Non avendo informazioni riguardo ad altri caratteri antropometrici oltre quelli sopra nominati, pensiamo di poter considerare i vari caratteri del cranio e della testa insieme con la lunghezza della testa; i caratteri che si riferiscono alla misura della lunghezza degli arti o di altre parti del corpo, pure composte da vari segmenti, insieme con la statura; i caratteri che si riferiscono a volumi (volume addominale, volume del tronco, volume toracico) o a diametri (diametro anteroposteriore toracico, diametro trasverso toracico) o alla superficie toracica e addominale, insieme col peso e col perimetro toracico.

Per i caratteri del primo tipo (testa e cranio) si hanno, su 145 indici per distribuzioni relative ad antichi abitanti adulti delle isole Canarie, 85 indici iponormali e 60 ipernormali; su 129 indici relativi ad altre popolazioni, 75 iponormali e 54 ipernormali. Si nota quindi, per questi caratteri, un maggior numero di distribuzioni iponormali rispetto a quelle ipernormali.

Per i caratteri del secondo tipo (statura e misure di lunghezza di arti e parti del corpo), su 50 distribuzioni esaminate 20 sono iponormali e 30 ipernormali.

Per i caratteri del terzo tipo (peso, volume, perimetro e diametro toracico, ecc.), su 15 distribuzioni 3 sono iponormali e 12 ipernormali.

In tutti e due questi ultimi tipi, il numero delle distribuzioni ipernormali supera quello delle distribuzioni iponormali, ma la percentuale delle distribuzioni ipernormali è molto più alta per i caratteri del terzo tipo che, come abbiamo detto dipendono più dall'ambiente, in confronto ai caratteri del secondo tipo.

Le nostre conclusioni risultano dunque confermate.

Le distribuzioni da noi considerate sono relative a popolazioni adulte.

Sarebbe stato preferibile poterci riferire a distribuzioni relative a coetanei, dato che, nelle distribuzioni complessive, vengono compresi dati relativi alle diverse età, sulle quali l'influenza dell'ambiente può essere diversa e accompagnata dall'influenza di molteplici altri fattori.

RIASSUNTO DELLA PRIMA PARTE.

6. Riassumiamo le conclusioni a cui siamo finora giunti.

Esaminate un gran numero di distribuzioni relative alla statura di giovani, in generale coscritti, abbiamo constatato che in quasi tutti i casi esse risultano ipernormali. Le distribuzioni relative alla statura di popolazioni complessive adulte, risultano invece in circa metà dei casi iponormali e in metà ipernormali.

Abbiamo esaminato poi distribuzioni relative a molteplici caratteri antropometrici, sia di coscritti, sia di popolazioni complessive adulte, e abbiamo constatato che le prime sono nel maggior numero dei casi ipernormali, le seconde in prevalenza iponormali.

Abbiamo quindi concluso che, nelle distribuzioni relative a popolazioni adulte, entra in azione un fattore, che tende ad abbassare o ad annullare la ipernormalità. Questo fattore potrebbe essere la variazione dei valori medi con l'età. La distribuzione totale verrebbe ad essere infatti la somma di più distribuzioni normali o ipernormali, con risultato, nel primo caso, che la distribuzione complessiva risulta iponormale e, nel secondo, che la ipernormalità della distribuzione complessiva risulta, se non negativa, in ogni modo minore di quella delle componenti.

Ci siamo poi chiesti se fosse possibile stabilire un'influenza dell'ambiente e dell'eredità sulla ipernormalità dei caratteri, ma

la nostra inchiesta non è esauriente per mancanza sia di dati precisi sull'influenza dell'ambiente e dell'eredità sui caratteri, sia di un numero sufficiente di distribuzioni da comparare relative a coetanei. In ogni modo, si è potuto notare che le distribuzioni dei caratteri più influenzati dell'ambiente risultano in generale più spesso ipernormali delle altre.

PARTE SECONDA

DATI SULLA RIFRAZIONE DELL'OCCHIO

7. Fra tutti i caratteri umani ve ne è uno, la rifrazione dell'occhio, che si stacca dagli altri in quanto presenta una fortissima ipernormalità. Essa è stata oggetto di studi dettagliati e approfonditi da parte degli oculisti, che li hanno eseguiti, non solo su popolazioni di razza caucasica d'Europa e degli Stati Uniti d'America, ma anche su altre stirpi, quali gli Eschimesi ed i Negri.

Dalle ricerche più estese e recenti che abbiamo trovato nell'abbondante letteratura in argomento, abbiamo ricavato alcune distribuzioni qui sotto elencate per le quali abbiamo determinato gli indici di ipernormalità.

(1) Dati della Clinica oculistica dell'Università di Tubinga, esposti in una memoria del 1929 dall'assistente Dott. Alwin Betsch (1).

Essi riguardano gli anni 1924, 1925, e 1927, e si riferiscono a 24.000 *pazienti*. L'esame però venne limitato ai pazienti sopra i 25 anni che non furono visitati per malattie della rifrazione, eliminando altresì i casi di astigmatismo, così che rimasero disponibili, per la statistica della rifrazione, 12.000 *occhi*.

L'indice d'ipernormalità, per il totale degli occhi, raggiunge il valore altissimo di 10,24.

Tra gli occhi esaminati, ve ne sono però alcuni che vengono riguardati come patologici perchè presentano anomalie nel loro fondo (principalmente occhi a cono). Essi rappresentano il 4%

(1) ALWIN BETSCH: *Über die menschliche Refraktionskurve*. «Klinische Monatsblätter für Augenheilkunde und für augenärztliche Fortbildung», Jahrgang 1929, 82. Band, pagg. 365-379.

dell'insieme degli occhi osservati e ad essi è dovuto il 27 % dei casi di miopia. Eliminati tali occhi come aberranti, l'indice di ipernormalità si riduce a 5,09, rimanendo in ogni modo molto alto.

Viceversa, la distribuzione degli occhi a cono, fortemente asimmetrica, risulta alquanto iponormale, con un indice di 3,075.

(2) Dati concernenti la clientela privata della Clinica oculistica dell'Università di Chicago, esposti in una memoria del 1931 dal prof. Peter C. Kronfeld e dalla Dottoressa Clarissa Devney, addetti alla clinica in parola (1).

Essi riguardano 10.000 pazienti di detta clinica di cui però furono considerati solo quelli trattati, ai fini della misurazione della rifrazione, con l'atropina. Furono inoltre scartati i pazienti sotto i 25 anni e sopra i 50, nonché quelli che presentavano un astigmatismo superiore a mezza diotria. Non pare siano stati invece esclusi gli occhi a cono. Vennero così considerati 2.229 occhi, che presentano un indice di ipernormalità = 5,20.

(3) Dati raccolti negli anni 1932 e 1935 a Göteborg in occasione delle leve militari, pubblicati nel 1936 dal Dott. E. Strömberg (2).

Riguardano 2.616 coscritti in età di 20 anni, di cui furono scartati 111 occhi per malattia, difetti di rifrazione, cataratta traumatica, anoftalmo o capacità visiva molto ridotta. Non pare che siano stati esclusi gli occhi a cono. L'esame risultò così limitato a 5.121 occhi. L'indice d'ipernormalità è = 8,83.

(4) Dati relativi a 1.000 individui (pazienti del Policlinico, medici, suore, allievi della scuola dei sottoufficiali, avieri), di cui 685 maschi e 315 femmine, pubblicati nel 1946 dal Dott. Sölve Stenström, addetto alla clinica oculistica di Uppsala (3).

Di ogni individuo venne esaminato solo un occhio. Vennero considerati anche i casi di astigmatismo, che spesso è accom-

(1) PETER C. KRONFELD und CLARISSA DEVNEY, *Ein Beitrag zur Kenntnis der Refraktionskurve*. «Albrecht von Graefe's Archiv für Ophthalmologie», Berlin, 126 Band, pagg. 487-501.

(2) E. STRÖMBERG, *Über Refraktion und Achselänge des menschlichen Auges*. «Acta Ophthalmologica», Vol. 14, Copenhagen, 1936.

(3) SÖLVE STENSTRÖM, *Untersuchungen über Variation und Kovariation der optischen Elemente des menschlichen Auges*. «Acta Ophthalmologica», Supplement XXI, Uppsala, 1946.

pagnato da anomalie della rifrazione. D'altra parte, all'infuori dei pazienti, la professione delle persone esaminate portava a considerare individui che non avevano potere visivo molto ridotto; quindi il materiale non comprende casi di forte miopia.

L'indice di ipernormalità risulta $= 6,24$ per i maschi, $= 4,24$ per le femmine, $= 5,75$ per il totale degli individui esaminati.

(5) Dati su 341 occhi di persone tra i 20 e i 60 anni che presentavano emmetropia o errori di rifrazione sferici, pubblicati nel 1957 dal Medical Research Council di Londra. ⁽¹⁾

L'indice di ipernormalità risulta uguale a 4,27.

È evidente che i risultati di queste cinque ricerche, relative a popolazioni europee non sono tra loro rigorosamente comparabili, il materiale risultando diversamente composto per età e per sesso, selezionato in vari sensi e i risultati differendo altresì per metodi di misurazione. Tutti i risultati ad ogni modo concordano nel mostrare una forte ipernormalità nella distribuzione della rifrazione.

Per le razze extra-europee, si posseggono dati per Negri e per Eschimesi.

(6) Dati su Negri del Gabon sono stati pubblicati nel 1937 da Stig Holm ⁽²⁾.

Comprendono maschi e femmine; vennero esclusi gli astigmatici. L'autore ha tenuto distinti gli analfabeti, in numero di 1.194, da coloro che sapevano leggere (in numero di 674, di cui 533 maschi e 141 femmine) e, tra questi, quelli che avevano letto per un tempo più o meno lungo, da quelli che avevano letto per poco tempo. I risultati ottenuti, senza escludere nessun caso, sono i seguenti:

Analfabeti: 1392 occhi di maschi, indice di ipernormalità $= 4,89$; 972 occhi di femmine, indice di ipernormalità $= 3,65$; complessivamente: 2364 occhi con un indice di ipernormalità $= 4,35$.

⁽¹⁾ ARNOLD SORSBY, B. BENJAMIN, J. B. DAVEY, M. SHERIDAN and J. M. TANNER. *Emmetropia and its Aberrations. A study in the correlation of the Optical Components of the Eye*. Medical Research Council, Special Report Series, No. 293, London 1957.

⁽²⁾ STIG HOLM, *Les états de la réfraction oculaire chez les Palénégrides au Gabon, Afrique Equatoriale Française. Étude de race pour éclairer la genèse de la refraction*, « Acta Ophthalmologica ». Supplementa 13-16, 1937-1939.

Persone che sapevano leggere: 1.310 occhi con un indice di ipernormalità = 6,07. Escludendo 8 occhi che eccedono 0,5 D di miopia e 3,25 D di ipermetropia, l'indice dei rimanenti 1.302 occhi, risulta = 3,36. Distinguendo secondo il tempo più o meno lungo durante il quale le persone hanno letto, si ottiene un indice di ipernormalità di 3,38 per 926 occhi di persone che hanno letto per un tempo più o meno lungo, e invece un indice di ipernormalità di 3,19 per 376 occhi di persone che hanno letto poco tempo.

(7) Dati sugli Eschimesi di Anmagssalik (costa orientale della Groenlandia) pubblicati nel 1954 da E. Skeller (1).

Essi comprendono 775 individui (382 maschi e 393 femmine), di tutte le età, da 2 a 69 anni. La distribuzione della rifrazione è data limitatamente agli occhi sferici, in numero di 1.923, eliminando quelli in cui l'astigmatismo supera mezza diotria. L'indice di ipernormalità risulta uguale a 3,74.

Casi di miopia non inferiore a 2 diotrie, non esistono fra gli Eschimesi e sono rari fra i Negri (5 casi su 2.364 occhi di analfabeti e pure 5 casi su 1.310 occhi di individui che sapevano leggere). Ciò spiega come la ipernormalità della distribuzione, a paragone di quella ottenuta per le altre distribuzioni, risulti bassa tra gli Eschimesi come tra i Negri.

SPIEGAZIONE DELLA IPERNORMALITÀ DELLE DISTRIBUZIONI CONCERNENTI LA RIFRAZIONE DELL'OCCHIO.

8. La forte ipernormalità della rifrazione dell'occhio non poteva non essere motivo di sorpresa, in quanto nulla di simile si conosceva per altri caratteri dell'uomo.

Non solo, ma anche le misure dell'occhio da cui la rifrazione dipende, sembravano avere una distribuzione normale o molto simile alla normale.

La rifrazione dell'occhio dipende infatti, da una parte, dal potere rifrangente dell'occhio, che a sua volta risulta dal potere rifrangente della cornea e dal potere rifrangente del cristallino, e, dall'altra, dalla lunghezza dell'asse dell'occhio. A sua volta il potere rifrangente del cristallino, collegato col suo spessore, dipende dalla

(1) *Anthropological and Ophthalmological Studies on the Angmagssalik Eskimos*, « Meddelelser om Groenlands », Kiobenhavn, Reitzel, 1954.

curvatura della sua superficie esterna e da quella della sua superficie interna, ossia dai rispettivi raggi, mentre la lunghezza dell'asse a sua volta dipende dalla profondità della camera anteriore, dallo spessore del cristallino e dalla profondità della camera posteriore. Ora, la profondità della camera anteriore, il raggio della cornea, lo spessore del cristallino, il raggio della sua superficie esterna come quello della superficie interna, il potere rifrangente del cristallino e il potere rifrangente della cornea e anche, quando venivano eliminati i casi estremi, la lunghezza dell'asse, presentavano distribuzioni che, tenuto conto del numero ristretto dei casi, venivano riguardate come normali. D'altra parte, l'aspettativa che anche le costanti dell'occhio, come le misure di tanti altri caratteri umani seguissero la curva normale faceva sì che si eliminassero come aberranti i casi estremi (per esempio, per la lunghezza dell'asse, i casi di forte miopia o di forte ipermetropia, oppure gli occhi a cono), procedimento arbitrario e ingannevole.

Risultando dalla combinazione di grandezze che si distribuivano normalmente, anche la rifrazione avrebbe dovuto presentare una distribuzione normale, qualora le varie componenti fossero state indipendenti nelle loro combinazioni. Si è però presto osservato che in realtà la dispersione della rifrazione era molto minore di quanto, in base ad una combinazione a caso delle sue componenti, sarebbe stato da attendersi e si dedusse quindi — e l'elaborazione statistica dei dati comprovò tale previsione — che le combinazioni non avvenivano a caso, ma che vi era un coordinamento tra le varie componenti, nel senso che le varianti delle diverse grandezze si compensavano tra loro così da determinare un addensarsi nei valori della rifrazione sulle gradazioni intermedie e quindi una variabilità della rifrazione minore di quanto sarebbe avvenuto nell'ipotesi di combinazioni a caso delle sue componenti.

Si credette così di avere spiegato la distribuzione ipernormale, in quanto questa è caratterizzata da una maggiore frequenza dei valori intermedi in confronto alla curva normale, ma in realtà la spiegazione era insoddisfacente perchè, oltre che da una frequenza maggiore dei valori intermedi, la curva ipernormale è caratterizzata, in confronto alla normale, da una frequenza maggiore dei valori estremi, che non si sarebbe dovuta verificare

per effetto della compensazione tra le grandezze che si combinavano.

La spiegazione si ebbe quando rilevazioni più estese, che prendevano debitamente in considerazione tutti i casi rilevati, senza eliminare gli estremi, misero in luce che anche la lunghezza dell'asse, componente importante della rifrazione, presentava una curva di frequenza ipernormale. Per illustrare questo punto, noi abbiamo calcolato gli indici di ipernormalità per le serie di Strömberg (Göteborg), di Stenström (Uppsala) e del Medical Research Council (Inghilterra), ottenendo rispettivamente i valori 4,05 ; 3,81 ; 3,92 ; tutti spiccatamente ipernormali.

Il calcolo, d'altra parte, del coefficiente di correlazione tra la lunghezza dell'asse e la rifrazione conferma la dipendenza di questa da quella. Per i dati di Uppsala, ad esempio, esso risulta $-0,76$.

I valori degli indici di ipernormalità della rifrazione risultano però, in ogni caso, notevolmente superiori a quelli della lunghezza dell'asse. Essi risultano invero rispettivamente, per le tre serie sopra considerate, uguali a 8.23 ; 5.75 ; 4.27.

Viene spontanea l'idea che la differenza sia collegata alla compensazione che si verifica tra le varie componenti della rifrazione, che ha per effetto di aumentare la frequenza dei valori intermedi di questa.

Ricerche recenti hanno infatti mostrato che, quando l'asse è molto lungo (più di 26,5 mm) o al contrario molto breve (meno di 22 mm), l'occhio risulta inevitabilmente miope o rispettivamente non miope ⁽¹⁾, senza che si verifichi una compensazione con le altre misure che valga a renderlo emmetrope.

I casi estremi per lunghezza dell'asse, che danno luogo ad una curva di frequenza ipernormale, rimangono dunque casi estremi anche nella curva di frequenza della rifrazione, contribuendo alla sua ipernormalità, mentre, all'interno di tali valori, si verifica la compensazione, che ha per effetto di addensare i valori della rifrazione intorno alle grandezze intermedie più di quanto non avvenga per i valori della lunghezza dell'asse, con la conseguenza che la ipernormalità della rifrazione risulta più accentuata di quella della lunghezza dell'asse.

(¹) Vedi STENSTRÖM, op. cit., pag. 73.

Si è quindi concluso che la ipernormalità delle curve della rifrazione, come quella della lunghezza dell'asse, dipende dalla eterogeneità del materiale.

Tale conclusione è esatta, ma la interpretazione, che di tale eterogeneità è stata data da qualche autore, è invece contestabile.

Gli è che si è scambiato quella che era solo una spiegazione sufficiente della ipernormalità con la sua spiegazione necessaria.

Per rendersi ragione della ipernormalità che si presenta nelle curve di frequenza di alcuni fenomeni, era stato invero osservato che, se si hanno due distribuzioni con la stessa media e con diversa variabilità, la distribuzione che risulta dalla loro somma è ipernormale, e di ciò può vedersi la dimostrazione nell'Appendice seconda.

Ma ciò non significa che, se una curva è ipernormale, ciò debba necessariamente avvenire perchè la distribuzione risulta dalla somma di due distribuzioni con la stessa media (o con medie simili) e con diversa variabilità.

Questa spiegazione viene tuttavia adottata, non so se da altri, ma certamente dallo Skeller, il quale, in un lavoro da molti punti di vista approfondito, ha anche ritenuto di trovare la conferma di tale tesi tracciando le curve cumulative di frequenza su carta di probabilità.

Se la distribuzione è normale, la supposta curva cumulativa assume la forma di una retta ; nel fatto, su tale carta, le curve anormali della rifrazione degli occhi assumono invece un andamento nettamente diverso, che si avvicina alla figura seguente. (fig. (8) ⁽¹⁾) la quale presenta una forte incurvatura nei punti A e B.

Lo Skeller riteneva che questa fosse appunto la forma che la curva della rifrazione avrebbe dovuto assumere qualora la sua distribuzione dipendesse dalla somma di due distribuzioni normali con medie approssimativamente uguali e con diversa variabilità. Ma ciò non è esatto ; la curva in parola assumerebbe in realtà in tale ipotesi l'aspetto della figura (9).

Si può invece ottenere un andamento approssimativamente corrispondente a quello della figura 8, ammettendo che la distri-

⁽¹⁾ ERIK SKELLER, nel libro sopracitato, ha disegnato le curve di probabilità corrispondenti ad alcune delle distribuzioni da me esaminate, relative alla rifrazione dell'occhio.

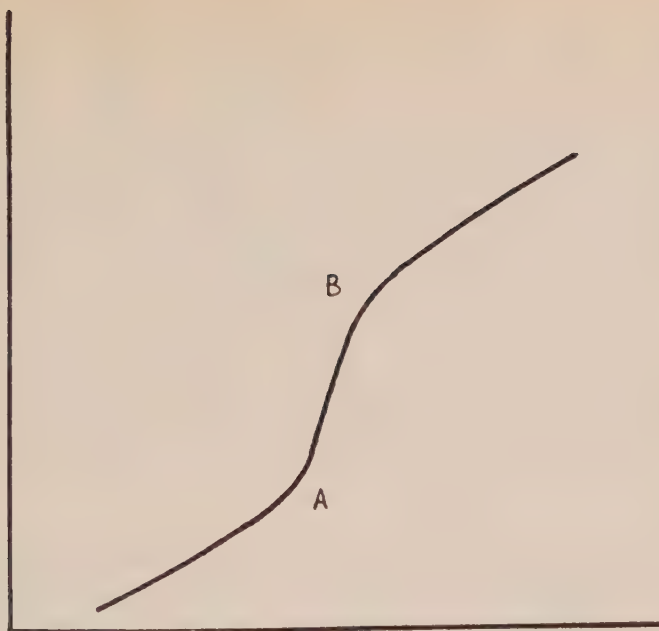


Fig. 8.

buzione ipernormale risulti dalla somma di tre distribuzioni normali, di cui una corrispondente ad un tipo medio prevalente e due corrispondenti a due ectipi, cioè, nel caso dell'occhio, a un tipo miope con media inferiore e a un tipo ipermetrope con media superiore al tipo prevalente.

L'idea che la distribuzione della rifrazione dell'occhio risulti dal sovrapporsi di tre distribuzioni corrispondenti a tre diversi gruppi biologici (occhi normali, miopi e ipermetropi) non è nuova. Essa si presenta anzi spontanea e, nel campo degli oculisti è stata svolta in particolare da Fredrik Berg ⁽¹⁾.

Il Berg si basa sul fatto che le correlazioni tra le varie misurazioni dell'occhio risultano diversamente alte per gli occhi fortemente ipermetropi (oltre $+ 2$ D) e, sia pure in grado minore, per gli occhi fortemente miopi (oltre $- 6$ D) che per gli inter-

⁽¹⁾ *Über Variabilität und Korrelation bei den verschiedenen Abmessungen des Auges*, «Gräfe's Archiv für Ophthalmologie» 127 Band, Berlin, 1931, pagg. 606-639.

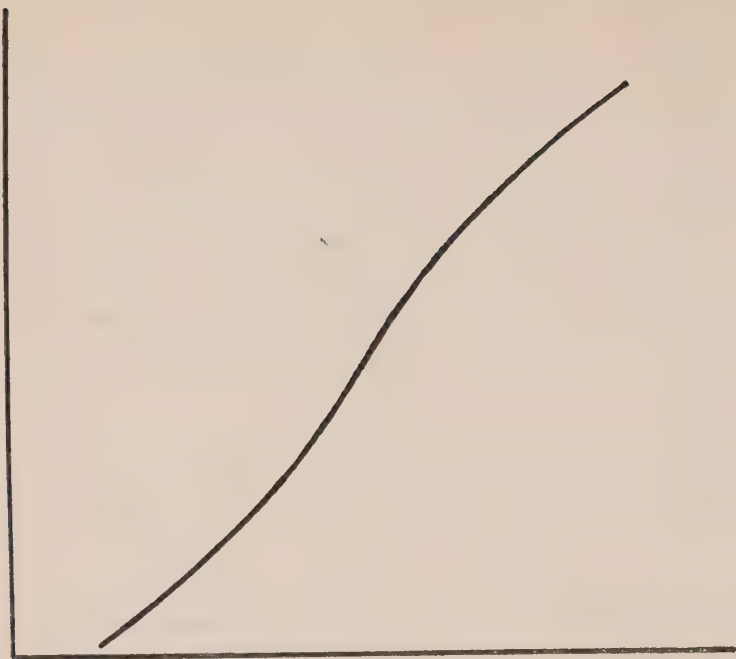


Fig. 9.

medi, per quanto naturalmente, come egli debitamente avverte, questi limiti non debbano essere scambiati con i limiti dei tre gruppi biologici, le cui misurazioni in diottrie, invece, ovviamente si devono sovrapporre. A conferma delle differenze biologiche fra i tre gruppi, egli porta l'osservazione, che sarebbe stata fatta, che la ipermetropia e la miopia, se moderate, parrebbero risultare dominanti nella ereditarietà, mentre, se sono molto forti, sarebbero recessive. In realtà, però, questo fatto potrebbe provare che la forte ametropia appartiene a un gruppo biologico diverso dalla ametropia lieve e dalla emmetropia, ma non che vi sono tre, piuttosto che due, gruppi biologici.

È da osservare a questo proposito che i costituzionalisti italiani (De Giovanni, Viola) parlavano della esistenza, nella conformazione del corpo umano, di un tipo e di due ectipi, ma tale intuizione non era stata accettata dagli statistici in quanto contrastava con l'esistenza di una distribuzione normale del carattere, quale si riteneva verificarsi. Essa è invece in accordo con

l'esistenza di una distribuzione ipernormale e in particolare con la forma che presenta la distribuzione della rifrazione dell'occhio umano.

Il fatto che, come abbiamo visto (pag. 53-54), la variabilità e l'ipernormalità della statura, del peso e del perimetro toracico sieno, nei coscritti inabili al servizio militare, spiccate in confronto a quelle degli individui normali e crescano con la gravità dei difetti degli individui considerati, può essere riguardato come una conferma di tale punto di vista dei costituzionalisti, in quanto più i difetti sono gravi e maggiore deve essere l'importanza degli ectipi in confronto al tipo normale.

Non sarebbe questa la prima volta che, approfondendo una questione, i tecnici giungono a confermare l'intuizione dei non tecnici, che, a tutta prima, l'esame dei tecnici sembrava dover escludere.

8 bis. Naturalmente non è detto che i due ectipi abbiano la stessa importanza numerica, nè che abbiano la stessa variabilità, nè che le loro distribuzioni sieno entrambe esattamente normali o, in ogni modo, simmetriche; non verificandosi tali ipotesi, la distribuzione risultante sarà asimmetrica, ed effettivamente asimmetriche sono in generale le distribuzioni della rifrazione dell'occhio.

Dal grafico di una distribuzione relativa all'indice di rifrazione degli occhi, effettuato su carta di probabilità, si vedono chiaramente i segmenti di retta, che corrispondono alle tre distribuzioni componenti; essi ci potrebbero dare anche delle indicazioni sulla relativa importanza di queste. Esaminiamo, per esempio, i dati esposti dal Dott. Alwin Betsch e riportiamo su carta di probabilità le curve cumulative della distribuzione totale e di quella ottenuta eliminando gli occhi a cono. Si hanno i grafici delle figure 10 e, rispettivamente, 11.

Si può notare che, nella prima curva, il segmento di retta che corrisponde al tipo miope è più lungo di quello che corrisponde al tipo ipermetrope, mentre, nella seconda curva, i due segmenti di retta sono quasi uguali.

Questo suggerisce che la distribuzione la quale corrisponde al tipo miope nella distribuzione globale o è più numerosa di quella che corrisponde al tipo ipermetrope oppure, a differenza di questa, è simmetrica o è, più di questa, asimmetrica. In real-

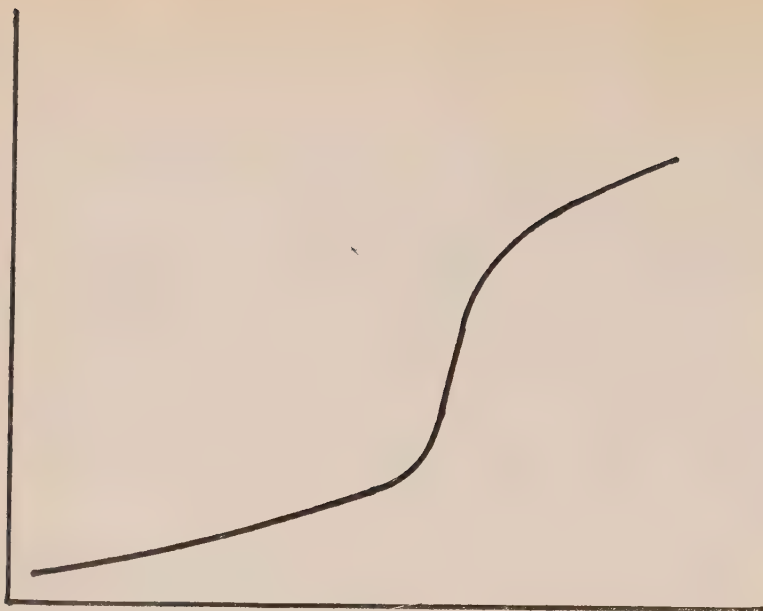


Fig. 10.

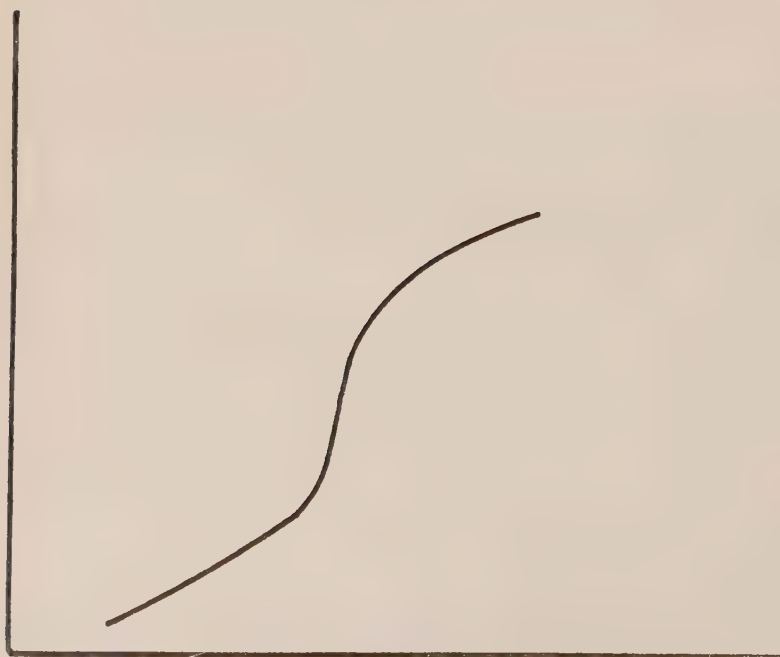


Fig. 11.

tà, tutte e due queste cause contribuiscono alla asimmetria della distribuzione globale. La distribuzione degli occhi a cono è infatti spiccatamente asimmetrica come si vede dal suo grafico effettuato su carta normale (fig. 12). Tali occhi a cono, oltre ad aumentare la numerosità degli occhi miopi in confronto agli occhi ipermetropi (numerosità che si riduce quasi uguale se si escludono i detti occhi a cono) rendono anche asimmetrica la distribuzione corrispondente al tipo miope.

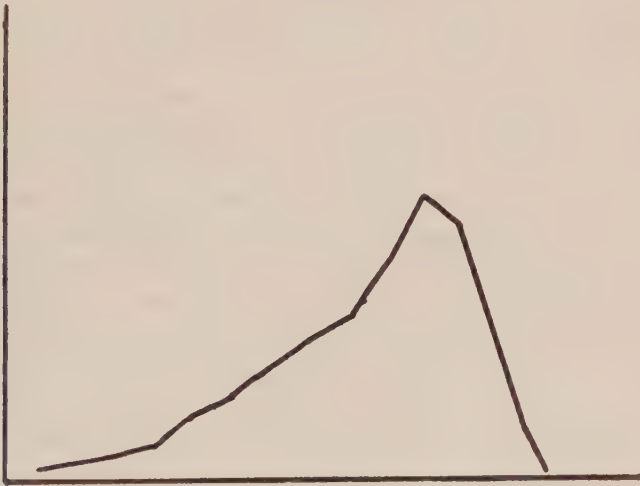


Fig. 12.

VARIE ALTRE SPIEGAZIONI DELLA IPERNORMALITÀ DELLE DISTRIBUZIONI DEI CARATTERI ANTROPOMETRICI.

9. È verosimile che la spiegazione dell'ipernormalità, che abbiamo suggerito e verificato per il caso della rifrazione dello occhio, valga anche per gli altri caratteri di coetanei che abbiamo passato precedentemente in rassegna (§ 2), ma, per questi, il grado di ipernormalità è troppo lieve perchè se ne possa eseguire la verifica con lo stesso procedimento.

Se essa vale, può fornire la spiegazione della diminuzione dell'ipernormalità che abbiamo riscontrato per le stature dei co-scritti italiani. Gli è che, col tempo, è diminuita l'importanza delle basse stature. Questa spiegazione si accorda con la contemporanea attenuazione che, nella variabilità e soprattutto nella asim-

metria delle distribuzioni, era stata osservata da S. Alberti ⁽¹⁾, da A. Costanzo ⁽²⁾, da L. Faleschini ⁽³⁾ e che a noi pure risulta dai calcoli eseguiti. Essa si può plausibilmente spiegare con le migliorate condizioni di vita delle popolazioni italiane nel periodo considerato. Se variazioni analoghe non si riscontrano per la Francia, ciò può forse attribuirsi al fatto che, durante il breve periodo considerato (1903-1922), intramezzato dalla grande guerra, non visi verificò un forte progresso.

Spiegazioni diverse, che il Gini si era prospettato, per spiegare la ipernormalità delle distribuzioni delle stature, non lo avevano lasciato soddisfatto ⁽⁴⁾.

Un'altra spiegazione, pure, a nostro modo di vedere, non soddisfacente, è stata prospettata dal Dott. Widukind Lenz ⁽¹⁾. Per il caso delle stature, egli ritiene che l'ipernormalità possa attribuirsi alla omogamia, cioè alla tendenza a sposarsi tra persone simili per statura. In confronto a quanto si verificherebbe in regime di panmixia (cioè se nessuna relazione si verificasse tra le stature dei due sposi), l'omogamia — osserva il Dott. Lenz — abbassa la variabilità del carattere. Ciò si può ammettere come tendenza, senza escludere però che gli effetti di questa siano compensati da altri fattori, per esempio dalla selezione; ma non si capisce perchè una diminuzione della variabilità debba far aumentare l'ipernormalità della distribuzione. Si ha l'impressione che il Dottor Lenz confonda la diminuzione della variabilità, che dipende dalla ampiezza della distribuzione e comporta un addensarsi dei valori intorno alla media, con l'aumento della ipernormalità, che dipende dalla forma della distribuzione e comporta, non solo un addensarsi dei casi intermedi, intorno alla media ma contemporaneamente un estendersi dei casi estremi al di là dei limiti teorici.

⁽¹⁾ *Influenza della diminuzione della mortalità*, Roma, Sormani, 1939, da A. Costanzo.

⁽²⁾ *La statura degli italiani ventenni nati dal 1854 al 1920*, « Annali di Statistica », serie VIII, vol. II 1948.

⁽³⁾ Vedi art. cit. a pag. 13.

⁽⁴⁾ Vedasi: *La distribuzione delle stature è ipernormale*, articolo citato pagg. 36-42.

⁽¹⁾ WIDUKIND LENZ, *Über den Einfluss der Homogamie auf die Verherbungskurven der menschlichen Körperhöhe*, « Acta Genetica et Statistica medica », vol. III, fasc. 2., 1952.

Il Dott. Lenz vede una conferma della sua ipotesi nel fatto che, in Italia, dal 1884 al 1920, l'indice di ipernormalità della statura risulta diminuito (vedi tavola II), mentre è da ritenere che si sia venuta attenuando, per l'intensificarsi delle comunicazioni, l'omogamia. Però in Francia, dal 1903 al 1922, non si è verificato alcuna attenuazione dell'indice di ipernormalità (vedi tavola VIII), per quanto sia da ritenere che, rispetto all'omogamia, la tendenza non abbia potuto essere diversa che in Italia.

D'altra parte, in Italia la variabilità delle stature non è affatto cresciuta, come, per effetto della ridotta omogamia, il Lenz si sarebbe atteso, ma al contrario diminuita.

Più elaborata spiegazione ha suggerito il Faleschini. Nell'articolo già citato, egli è giunto alla conclusione che l'indice di ipernormalità di una distribuzione complessiva risultante dalla somma di più distribuzioni gaussiane (la sua dimostrazione si riferisce propriamente all'indice di Pearson, ma la conclusione può considerarsi come avente valore generale), è la risultante della somma di tre addendi: 1°) l'indice di ipernormalità delle medie delle distribuzioni parziali dalla cui somma risulta la distribuzione complessiva; 2°) la varianza di dette medie; 3°) la covarianza tra gli scostamenti delle medie delle distribuzioni parziali dalla media della distribuzione complessiva, e le varianze delle dette distribuzioni parziali.

Il valore probabile del 3° addendo può ritenersi nullo; quello del 2° è ovviamente sempre positivo; quello del 1° è ritenuto dal Faleschini avere un valore normale o ipernormale. Se così veramente fosse, è ovvio che l'indice di ipernormalità della distribuzione complessiva sarebbe necessariamente ipernormale.

Ma perchè l'indice delle medie delle distribuzioni parziali non potrebbe essere iponormale? Perchè, ad esempio, non potrebbe essere iponormale la distribuzione delle stature medie delle varie età? Il fatto, da noi messo in luce, che i coetanei presentano generalmente distribuzioni ipernormali delle stature e invece gli adulti di tutte le età molto spesso distribuzioni normali o iponormali fa pensare proprio che si verifichi tale ipotesi.

E, analogamente, perchè le distribuzioni delle stature medie delle varie circoscrizioni di un territorio o dei vari anni di un più lungo periodo di tempo non potrebbero presentare indici iponormali? Il fatto, da noi constatato, che gli indici di ipernor-

malità delle stature per l'intero territorio e la media degli indici di ipernormalità per le singole circoscrizioni non differiscano sistematicamente fa pensare anche in questo caso a detta possibilità. Qualche riscontro, che ho eseguito, conferma perfettamente tale aspettativa.

La tavola XXXIII-*bis* contiene le medie delle stature per i vari compartimenti italiani del 1915, che hanno servito al calcolo

TAVOLA XXXIII *bis**Stature medie nei compartimenti italiani nel 1915*

Compartimenti	Medie	Compartimenti	Medie
Piemonte	167,459	Umbria	164,940
Liguria	167,916	Lazio	165,800
Lombardia.	166,842	Abruzzi e Molise	163,431
Venezia Tridentina	167,107	Campania	163,812
Veneto	167,994	Puglie.	162,870
Venezia Giulia e Zara.	169,840	Lucania	161,649
Emilia.	166,896	Calabria	163,402
Toscana	167,366	Sicilia	163,452
Marche	164,668	Sardegna.	161,375

MEDIA DELLE MEDIE = 165,379

degli indici di ipernormalità esposti nella terz'ultima colonna della tavola II, e la tavola XXXIII-*ter* le medie delle stature delle provincie lombarde e toscane, pure nel 1915, da cui sono stati ricavati gli indici di ipernormalità esposti alla tavola IV. L'indice di ipernormalità per la distribuzione della tav. XXXIII-*bis* risulta = 2,540 ; quello per la distribuzione della tav. XXXIII-*ter*

TAVOLA XXXIII-ter

Stature medie nelle province lombarde e toscane (1915)

LOMBARDIA		TOSCANA	
Province	Medie	Province	Medie
Bergamo.	166,362	Arezzo.	165,527
Brescia	165,679	Firenze	168,058
Como	166,934	Grosseto.	165,642
Cremona.	166,019	Livorno	168,206
Mantova.	166,820	Lucca	168,642
Milano.	167,238	Massa Carrara	166,748
Pavia	167,571	Pisa.	168,348
Sondrio	166,353	Pistoia.	168,109
Varese.	168,042	Siena	165,502
MEDIA DELLE MEDIE . . .	166,780	MEDIA DELLE MEDIE . . .	167,198

relativo alle province della Toscana, = 2,226 e quello relativo alle province della Lombardia, = 2.832.

La tavola XXXIII-*quater* contiene le medie relative alle province e circondari della Lombardia e la tavola XXXIII-*quinquies* quelle relative alle province e circondari della Toscana per il 1898, che hanno servito al calcolo degli indici di ipernormalità contenuti nelle tavole V e VI. Ora gli indici di ipernormalità ricavati dalla tav. XXXIII-*quater* sono : 2.522 per le province e 3,380 per i circondari ; quelli relativi alla tavola XXXIII-*quinquies* sono 2,460 per le province e 2,636 per i circondari.

Concludendo, su sette casi, gli indici per le distribuzioni delle medie risultano, in sei, iponormali e in un solo caso ipernormali.

TAVOLA XXXIII-quater

Stature medie nelle province e nei circondari della Lombardia (1898)

Province	Medie	Circondari	Medie
Bergamo.	163,554	Bergamo. 164,270 Clusone 160,867 Treviglio. 163,621	
Brescia	163,013	Breno 162,382 Brescia 163,231 Chiari 161,596 Salò. 165,473 Verolanuova 162,347	
Como	164,180	Como 164,132 Lecco 162,692 Varese. 165,623	
Cremona.	163,720	Casalmaggiore 163,801 Crema. 163,388 Cremona. 163,883	
Mantova.	164,365	Mantova. 164,365	
Milano.	164,603	Abbiategrasso 164,631 Gallarate. 164,854 Lodi. 164,581 Milano. 164,968 Monza. 163,536	
Pavia	164,628	Bobbio 163,761 Lomellina 165,340 Pavia 164,514 Voghera 164,218	
Sondrio	163,702	Sondrio 163,702	
MEDIA DELLE MEDIE . . .	163,971	MEDIA DELLE MEDIE . . .	163,831

TAVOLA XXXIII-*quinquies**Stature medie nelle province e nei circondari della Toscana (1898)*

Province	Medie	Circondari	Medie
Arezzo.	164,237	Arezzo.	164,088
Firenze	165,471	Firenze 165,599 Pistoia. 165,962 Rocca S. Casciano 161,699 S. Miniato. 165,841	
Grosseto.	163,141	Grosseto.	163,141
Livorno	166,906	Isola d'Elba 164,917 Livorno 167,309	
Lucca	166,385	Lucca	166,385
Massa e Carrara	164,450	Castelnuovo 166,220 Massa 164,121 Pontremoli. 163,605	
Pisa.	165,711	Pisa. 166,475 Volterra 163,490	
Siena	163,604	Montepulciano 162,250 Siena 164,325	
MEDIA DELLE MEDIE . . .	164,988	MEDIA DELLE MEDIE . . .	164,714

IPERNORMALITÀ E VARIABILITÀ DEI CARATTERI STRUTTURALI
E DEI CARATTERI FUNZIONALI

La rifrazione dell'occhio è un carattere funzionale; tutti i caratteri a cui si riferiscono gli altri indici di ipernormalità considerati finora in questo articolo sono, invece, caratteri strutturali. È possibile — dobbiamo domandarci — che da ciò dipenda la forte ipernormalità che abbiamo riscontrato per la rifrazione dell'occhio in confronto a quella di tutti gli altri caratteri?

La risposta pare debba essere affermativa. Si è visto infatti che la forte ipernormalità della rifrazione dell'occhio si può spiegare con la compensazione che si verifica tra le intensità dei vari caratteri strutturali che concorrono a determinare detta rifrazione, fin tanto che essi si mantengono nei limiti normali del

tipo, mentre la compensazione non si verifica più quando la lunghezza dell'asse eccede tali limiti e appartiene ad uno dei due ectipi.

Nulla di simile sembra di regola avvenire per i caratteri strutturali, o quanto meno per quelli lineari, che si possono riguardare come il risultato della somma di vari segmenti giustapposti, in quanto, tra i segmenti che li compongono, in generale non si verifica compensazione e neanche indipendenza, ma invece un certo grado di solidarietà.

Sarà interessante ricercare se i caratteri funzionali (o alcuni tra essi) presentano una ipernormalità particolarmente elevata in confronto ai caratteri strutturali.

Ho avuto occasione, in altro gruppo di ricerche ⁽¹⁾, di calcolare, per 8 serie di dati ricavati da Farris ⁽²⁾, gli indici di ipernormalità per la durata del ciclo sessuale della donna e per le sue due componenti: il periodo pre-ovulare e il periodo post-ovulare. Essi sono riportati nella tavola XXXIV.

La media degli 8 indici di ipernormalità risulta uguale a 3,50 per il periodo pre-ovulare, a 4,49 per il periodo post-ovulare e a 4,64 per la lunghezza totale del ciclo sessuale. Come si vede, questi valori sono notevolmente superiori a quelli trovati per i caratteri strutturali e il valore per la lunghezza del ciclo sessuale è dello stesso ordine di grandezza di alcuni indici di ipernormalità ottenuti per la rifrazione dell'occhio.

Si nota anche che l'indice di ipernormalità per la lunghezza del ciclo sessuale è sensibilmente superiore agli indici delle due parti componenti, e ciò vale, non solo per le medie sopra indicate, ma anche per gli indici di ipernormalità relativi alle singole serie. Questa differenza può forse ammettere una spiegazione analoga a quella che abbiamo dato per la superiorità dell'indice di ipernormalità relativo alla rifrazione dell'occhio in confronto agli indici di ipernormalità relativi ai caratteri che lo determinano.

In generale, gli indici di ipernormalità risultano più elevati per le serie che presentano una maggiore variabilità (desunta

⁽¹⁾ Si tratta di ricerche statistiche sul ciclo sessuale della donna, di cui mi riservo di pubblicare prossimamente i risultati di ricerche che sono ultimate, ma non ancora stesi.

⁽²⁾ E. J. FARRIS, *Human Ovulation and Fertility*, J.B. Lippincott, Philadelphia. 1956.

TAVOLA XXXIV

Scostamento semplice medio (1S), scostamento quadratico medio (2S) e indice di ipernormalità (I) per la durata del ciclo sessuale e delle sue componenti (periodo preovulare e periodo postovulare).

NUMERO DELLE OSSERVAZIONI ESEGUITE	CICLO SESSUALE			PERIODO PREOVULARE			PERIODO POST-OVULARE			OSSERVAZIONI
	1S	2S	I	1S	2S	I	1S	2S	I	Pagine del volume di Farris in cui sono pubblicati i dati
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
144	3,03	4,27	3,96	1,88	2,37	3,18	2,52	3,44	3,74	23
100	2,19	3,05	3,90	1,71	2,33	3,71	1,71	2,22	3,38	24
84	2,43	3,82	4,97	1,66	2,18	3,45	1,78	2,59	4,20	46
150	1,96	2,55	3,40	1,59	2,00	3,20	1,58	2,07	3,44	48
30	3,14	6,50	8,58	1,79	2,67	4,46	2,41	5,29	9,61	52
100	2,03	2,69	3,49	1,43	1,84	3,32	1,60	2,04	3,27	58
364	2,19	3,42	4,86	1,58	2,07	3,44	1,70	2,62	4,74	67
761	2,21	3,12	3,98	1,58	2,01	3,23	1,81	2,44	3,54	71

Indice di cograduazione fra i dati delle colonne :

$$3 \text{ e } 1 = 0,75$$

$$6 \text{ e } 4 = 0,125$$

$$9 \text{ e } 7 = 0,375$$

$$3 \text{ e } 2 = 0,75$$

$$6 \text{ e } 5 = 0,437$$

$$9 \text{ e } 8 = 0,813$$

dallo scostamento quadratico medio o dallo scostamento semplice medio), come si vede dagli indici di cograduazione, tutti positivi e di valore elevato, indicati in calce alla tavola. Varrà la pena di riscontrare se questa relazione abbia carattere generale.

Varrà pure la pena di riscontrare le relazioni tra gli indici di ipernormalità del carattere risultante e dei caratteri componenti, sia per i caratteri funzionali che per i caratteri strutturali.

Per il ciclo sessuale, il valore dell'indice di cograduazione tra i valori della colonna 3 e quelli della colonna 6 è $= 0.438$ e quello dell'indice di cograduazione tra i valori della colonna 3 e quelli della colonna 9 è $= 0.688$.

In una serie di ricerche relative alla variabilità delle ossa (metacarpo e falangi) della mano e del piede (caratteri struttu-

TAVOLA XXXV.

Indici di ipernormalità per le ossa della mano dei due sessi.

I = Pollice; II = Indice; III = Medio; IV = Anulare; V = Mignolo; M = Metacarpo;
 $F_1 = 1$ Falange; $F_2 = 2$ Falange; $F_3 = 3$ Falange; $D = F_1 + F_2 + F_3$ = Dito
 $R = M + D$ = raggio

UOMO, MANO DESTRA

DITO	M	F_1	F_2	F_3	D	R	DITO
I	3.55	3.50	—	3.42	3.33	3.48	I
II	3.09	3.05	3.34	2.81	3.30	3.36	II
III	3.33	3.36	3.10	2.95	3.24	3.31	III
IV	3.21	3.22	3.06	3.95	3.28	3.43	IV
V	3.25	3.26	3.58	4.00	3.32	3.33	V
Medie delle 5 dita .	3.29	3.28	3.27	3.42	3.29	3.38	

UOMO, MANO SINISTRA

DITO	M	F_1	F_2	F_3	D	R	DITO
I	3.12	3.71	—	3.19	3.15	3.58	I
II	3.19	3.13	3.19	3.08	3.02	3.16	II
III	3.33	3.28	3.11	2.95	3.17	3.68	III
IV	3.35	3.25	3.25	2.96	3.27	3.12	IV
V	3.30	3.30	3.36	3.21	3.45	3.20	V
Medie delle 5 dita .	3.32	3.34	3.23	3.08	3.21	3.35	

DONNA, MANO DESTRA

DITO	M	F_1	F_2	F_3	D	R	DITO
I	3.55	2.90	—	3.04	2.75	2.00	I
II	3.45	2.97	3.46	3.71	2.00	3.26	II
III	3.35	2.88	3.26	2.95	2.95	3.09	III
IV	3.19	3.03	3.10	3.22	3.13	3.11	IV
V	3.44	2.83	3.06	2.00	3.21	2.94	V
Medie delle 5 dita .	3.40	2.92	3.22	2.98	2.81	2.88	

DONNA, MANO SINISTRA

DITO	M	F_1	F_2	F_3	D	R	DITO
I	3.56	2.50	—	3.17	3.45	3.36	I
II	3.30	2.93	2.98	3.50	3.03	3.07	II
III	3.23	2.91	3.24	3.01	2.77	2.91	III
IV	3.52	2.81	3.06	3.33	3.90	2.87	IV
V	3.05	2.76	3.16	2.81	3.10	2.91	V
Medie delle 5 dita .	3.13	2.78	3.11	3.16	3.25	3.03	

TAVOLA XXXVI

Indici di ipernormalità per le ossa del piede dei due sessi.

I = Pollice; II = Indice; III = Medio; IV = Anulare; V = Mignolo; M = Metatarso;
 $F_1 = 1$ Falange; $F_2 = 2$ Falange; $F_3 = 3$ Falange; D = $F_1 + F_2 + F_3$ = Dito
 $R = M + D$ = raggio

UOMO, PIEDE DESTRO

DITO	M	F_1	F_2	F_3	D	R	Dito
I	3.18	3.33	—	3.08	3.27	3.00	I
II	3.15	3.16	3.14	3.40	3.43	3.33	II
III	3.29	3.24	2.53	2.52	3.06	3.32	III
IV	3.20	2.90	2.75	2.90	3.15	3.16	IV
V	3.20	3.18	3.78	3.12	3.16	3.13	V
Medie delle 5 dita .	3.20	3.16	3.05	3.20	3.22	3.19	

UOMO, PIEDE SINISTRO

DITO	M	F_1	F_2	F_3	D	R	Dito
I	3.27	3.22	—	3.01	3.20	3.11	I
II	3.33	3.29	3.07	3.47	3.33	3.19	II
III	3.24	2.92	3.63	3.40	3.16	3.14	III
IV	3.42	2.89	2.92	2.99	3.06	3.13	IV
V	3.00	2.90	3.62	2.99	3.24	3.03	V
Medie delle 5 dita .	3.25	3.04	3.06	3.17	3.20	3.12	

DONNA, PIEDE DESTRO

DITO	M	F_1	F_2	F_3	D	R	Dito
I	4.16	3.40	—	3.68	3.24	3.80	I
II	3.45	3.47	3.11	2.82	2.97	3.25	II
III	3.42	3.14	2.84	2.95	3.38	3.42	III
IV	3.34	2.96	4.29	2.96	3.08	3.35	IV
V	3.70	3.24	4.71	3.58	3.58	3.41	V
Medie delle 5 dita .	3.61	3.24	3.74	3.20	3.25	3.25	

DONNA, PIEDE SINISTRO

DITO	M	F_1	F_2	F_3	D	R	Dito
I	3.94	3.02	—	3.23	2.95	3.70	I
II	3.60	3.53	3.17	2.95	3.48	3.53	II
III	3.40	3.10	2.81	3.03	3.34	3.44	III
IV	3.27	3.29	2.00	2.97	3.54	3.59	IV
V	3.62	3.54	4.34	3.92	3.47	3.79	V
Medie delle 5 dita .	3.57	3.30	3.08	3.22	3.36	3.61	

rali), ho avuto occasione di calcolare, sui dati di Pfitzner ⁽¹⁾, gli indici di ipernormalità (vedi tavole XXXV e XXXVI). Per lo più (in quasi due terzi dei casi), questi risultano superiori al valore normale (3, 1416), ma la ipernormalità risulta leggera (in media, l'indice è = 3,22), come per la gran parte dei caratteri strutturali considerati in questo articolo. Non appare una superiorità della ipernormalità del dito in confronto a quella delle falangi da cui esso è composto, nè del raggio in confronto al dito e al metacarpo di cui è la somma, in contrasto con quanto si era rilevato per le relazioni tra la ipernormalità della rifrazione dell'occhio o del ciclo sessuale (caratteri funzionali) e quelle delle rispettive componenti e in accordo con quanto è risultato per le relazioni tra la ipernormalità della statura di una popolazione e quella dei gruppi (distinti secondo il territorio, o l'età, o l'anno di rilevazione) che la compongono.

Qualche confronto permettono anche i dati relativi agli iscritti di leva della Svizzera, raccolti dallo Schlaginhaufen: la media degli indici di ipernormalità relativi alla statura ha infatti un valore di 3,278, mentre le medie degli indici delle due componenti (altezza seduto e lunghezza della gamba) sono entrambe alquanto superiori (rispettivamente 3,343 e 3,306).

Ulteriori confronti possono essere fatti, sia per i padri che per le madri, tra gli indici di ipernormalità delle stature da una parte, e quelli delle sue componenti (lunghezza della faccia, lunghezza del tronco, lunghezza degli arti inferiori) dall'altra, nonché tra gli indici di ipernormalità della lunghezza del tronco e quelli delle sue componenti (lunghezza sternale, lunghezza xifo-epigastrica e lunghezza epigastrico-pubica) nelle famiglie numerose, del Comune di Pisino (Istria). Sia per i padri che per le madri, sia per la statura che per la lunghezza del tronco, gli indici di ipernormalità risultano intermedi (e non superiori) a quelli per le loro componenti ⁽²⁾.

⁽¹⁾ W. PFITZNER, *Beiträge zur Kenntniss des menschlichen Extremitätenskelets* «Morphologische Arbeiten, herausgegeben von Dr. Gustav Schwalbe» Erster Band, Erstes Heft, 1892; Zweiter Band, Erstes Heft, 1893. Jena, G. Fischer.

⁽²⁾ Le misurazioni vennero eseguite dal Dr. GIOVANNI PESANTE su tutti i genitori con sette o più figli del Comune di Pisino in occasione di una indagine antropometrica organizzata nel 1931 dal Comitato Italiano

La tavola XXXVII contiene, oltre agli indici di ipernormalità, gli indici di variabilità desunti dallo scostamento semplice medio percentuale ($100\sigma S/A$) e dallo scostamento quadratico medio percentuale (ragguagliato alla media aritmetica) per questi e per alcuni altri caratteri.

TAVOLA XXXVII

Scostamento semplice medio percentuale ($100\sigma S/A$), scostamento quadratico medio percentuale ($100\sigma^2 S/A$) e indice di ipernormalità (I) per i caratteri antropometrici dei padri (in numero di 1019) e delle madri (in numero di 1191) di famiglie numerose del Comune di Pisino (Istria).

CARATTERI ANTROPOMETRICI	PADRI			MADRI		
	$100\sigma S/A$	$100\sigma^2 S/A$	I	$100\sigma S/A$	$100\sigma^2 S/A$	I
	(1)	(2)	(3)	(4)	(5)	(6)
Statura in piedi	3,03	3,96	3,41	3,28	4,18	3,25
Apertura delle braccia	3,23	4,14	3,28	3,24	4,05	3,12
Lunghezza della faccia	5,23	7,15	4,04	5,13	7,09	3,82
Lunghezza sternale	6,77	8,77	3,35	7,01	9,38	3,58
Lunghezza xifo-epigastrica	13,75	17,11	3,10	13,06	16,78	3,30
Lunghezza epigastrico-pubica	6,46	8,35	3,34	7,05	8,98	3,24
Lunghezza del tronco	5,52	7,06	3,27	6,16	7,77	3,18
Lunghezza dell'arto inferiore	3,51	4,90	3,90	4,16	5,40	3,36
Statura seduta	2,55	3,19	3,12	2,57	3,26	3,23

Indice di cograduazione fra i dati delle colonne :

$$3 \text{ e } 1 = -0,15$$

$$6 \text{ e } 4 = +0,25$$

$$3 \text{ e } 2 = -0,05$$

$$6 \text{ e } 5 = +0,30$$

per lo studio dei problemi della Popolazione (per le relazioni sommarie dei risultati di detta indagine vedi il volume IV, Sezione di Antropologia e Geografia, degli Atti del Congresso internazionale degli studi sulla popolazione, pubblicati dal detto Comitato, Roma, 1934). I dati individuali esistenti nell'archivio del Comitato vennero elaborati dallo studente Salvatore Matricardi per la sua tesi di laurea presentata alla Facoltà di Scienze Politiche dell'Università di Roma dal titolo *La variabilità della statura e dei segmenti che la compongono*. Il Matricardi ha anche calcolato le medie, gli scostamenti semplici e quadratici per i nove caratteri a cui si riferisce la tavola. XXXVII. I valori per gli scostamenti semplici e quadratici, indicati nella tavola sono quelli calcolati in base a classi di 0,5 cm. per la statura, l'apertura delle braccia, la lunghezza dell'arto inferiore e la statura seduta e per classi di 0,1 cm. per gli altri caratteri.

In calce alla tavola sono indicati gli indici di cograduazione tra gli indici di ipernormalità e gli indici di variabilità. Si nota che detti indici di cograduazione risultano diversi per i maschi e per le femmine, lievemente negativi per quelli, sensibilmente positivi per queste.

TAVOLA XXXVIII

Scostamento semplice medio relativo ($100 \frac{1S}{\sqrt{A}}$), scostamento quadratico relativo ($100 \frac{2S}{\sqrt{A}}$) e indice di ipernormalità (I) per i caratteri antropometrici dei padri (in numero di 1019) e delle madri (in numero di 1191) di famiglie numerose del Comune di Pisino (Istria)

CARATTERI ANTROPOMETRICI	PADRI			MADRI		
	$100 \frac{1S}{\sqrt{A}}$	$100 \frac{2S}{\sqrt{A}}$	I	$100 \frac{1S}{\sqrt{A}}$	$100 \frac{2S}{\sqrt{A}}$	I
	(1)	(2)	(3)	(4)	(5)	(6)
Statura in piedi	39,2	51,2	3,41	40,7	51,9	3,25
Apertura delle braccia	41,9	53,7	3,28	40,2	50,2	3,12
Lunghezza della faccia	21,3	30,2	4,04	21,4	29,5	3,82
Lunghezza sternale	31,4	40,7	3,35	31,2	41,8	3,58
Lunghezza xifo-epigastrica	49,2	61,2	3,10	42,3	54,3	3,30
Lunghezza epigastrico-pubica	32,9	42,5	3,34	34,8	44,3	3,24
Lunghezza del tronco	42,9	54,0	3,27	45,6	57,5	3,18
Lunghezza dell'arto inferiore	31,1	43,4	3,90	35,1	45,5	3,36
Statura seduta	29,0	36,2	3,12	28,6	36,3	3,23

Indice di cograduazione tra le colonne :

$$3 \text{ e } 1 = -0,50$$

$$6 \text{ e } 4 = -0,30$$

$$3 \text{ e } 2 = -0,40$$

$$6 \text{ e } 5 = -0,30$$

È però discutibile se, ai fini di eliminare l'influenza della diversa dimensione del carattere, sia appropriato dividere lo scostamento semplice medio $1S$ e lo scostamento quadratico medio $2S$ per la media aritmetica del carattere. Può parere preferibile dividerli per le radici quadrate dei detti indici. In tal caso, gli indici di variabilità relativa sono quelli indicati nella tavola XXXVIII e gli indici di cograduazione scritti in calce risultano,

sia per il padre che per la madre, tutti nettamente negativi e abbastanza elevati.

Per la lunghezza del ciclo sessuale e delle sue componenti, abbiamo trovato invece indici di cograduazione positivi (vedi pag. 83).

È da domandarsi se questi dati attestino, anche in questo campo, la possibilità di un diverso comportamento dei caratteri antropometrici strutturali e dei caratteri funzionali.

I caratteri strutturali considerati alle tabole XXXVII e XXXVIII sono tutti caratteri lineari e a caratteri lineari si riferisce il ragionamento fatto a pag. 82 che spiega la minore ipernormalità delle loro distribuzioni in confronto a quelle dei caratteri funzionali.

Resta da esaminare come si comportino i caratteri strutturali non lineari, che possono risultare, non dalla giustapposizione, ma dalla combinazione di caratteri strutturali lineari; per esempio, la circonferenza che è funzione del prodotto dei due diametri di lunghezza e larghezza; il volume che è funzione dei tre diametri di lunghezza, larghezza ed altezza.

La tavola XXXIX contiene i dati, relativi alla lunghezza, larghezza e circonferenza, per cinque gruppi di crani delle isole Canarie (maschi di Teneriffa; femmine di Teneriffa; maschi di Gomera; maschi delle quattro isole Gran Canaria, Fuerteventura, La Palma, Hierro; maschi di tutte e sei le predette isole).

In un caso (maschi di Gomera), il valore dell'indice di ipernormalità per la circonferenza, carattere risultante, è intermedio tra quelli dei due caratteri componenti (lunghezza e larghezza); in due casi (femmine di Teneriffa, maschi delle quattro isole) è superiore; in due (maschi di Teneriffa, maschi delle sei isole) è inferiore. Questo comportamento è simile a quello dei caratteri strutturali lineari, non a quello dei caratteri funzionali, in cui il valore dell'indice di ipernormalità è, per il carattere risultante, più elevato che per i caratteri componenti.

Per ciò che concerne la relazione tra ipernormalità e variabilità delle distribuzioni, gli indici di cograduazione tra i quindici valori degli indici di ipernormalità (col. 5) e i corrispondenti valori degli indici di variabilità, sia percentuali (col. 1 e 2) che relativi (col. 3 e 4), risultano tutti positivi, come abbiamo riscontrato avvenire per i caratteri funzionali.

TAVOLA XXXIX

Scostamento semplice medio percentuale ($100 \sqrt[1]{S/A}$) e relativo ($100 \sqrt[1]{S/\sqrt{A}}$), scostamento quadratico medio percentuale ($100 \sqrt[2]{S/A}$) e relativo ($100 \sqrt[2]{S/\sqrt{A}}$) e indice di ipernormalità (I) per la lunghezza, la larghezza e la circonferenza del cranio di sei delle isole Canarie (Teneriffa, Gomera, Fuerteventura, Gran Canaria, La Palma, Hierro)

ISOLE	SESSO DEGLI INDIVIDUI	CARATTERE DEL CRANIO	$100 \sqrt[1]{S/A}$ (1)	$100 \sqrt[2]{S/A}$ (2)	$100 \sqrt[1]{S/\sqrt{A}}$ (3)	$100 \sqrt[2]{S/\sqrt{A}}$ (4)	I (5)
Teneriffa	Maschi . .	Lunghezza	2,34	2,91	31,9	39,7	3,094
		Larghezza	2,83	3,53	33,7	42,1	3,116
		Circonferenza . . .	2,11	2,60	48,2	59,5	3,052
Teneriffa	Femmine .	Lunghezza	2,21	2,63	29,3	34,9	2,842
		Larghezza	1,97	2,43	23,1	28,5	3,046
		Circonferenza . . .	1,85	2,29	41,4	51,2	3,052
Gomera	Maschi . .	Lunghezza	2,85	3,45	33,4	46,4	2,920
		Larghezza	2,86	3,49	33,9	41,4	2,978
		Circonferenza . . .	2,18	2,66	49,4	60,2	2,968
Fuerteventura . .	Maschi . .	Lunghezza	3,33	4,00	45,3	54,5	2,862
Gran Canaria . . .		Larghezza	3,06	3,83	36,2	45,3	3,122
La Palma Hierro .		Circonferenza . . .	2,44	3,07	55,5	69,6	3,148
Tutte le sei isole .	Maschi . .	Lunghezza	2,55	3,19	34,8	43,5	3,128
		Larghezza	2,88	3,59	34,3	42,8	3,106
		Circonferenza . . .	2,22	2,74	50,6	62,5	3,048

Indice di cograduazione tra i dati delle colonne :

$$5 \text{ e } 1 = 0,18$$

$$5 \text{ e } 3 = 0,18$$

$$5 \text{ e } 2 = 0,13$$

$$5 \text{ e } 4 = 0,04$$

È però da tener presente che, nella tavola XXXIV, si erano considerati gli stessi caratteri per gruppi diversi di individui, mentre, nelle tavole XXXVII e XXXVIII, si sono considerati caratteri diversi per lo stesso gruppo di individui. Ora, è da domandarsi se il diverso comportamento degli indici di cograduazione non dipenda da questa differenza, anzichè dalla circostanza che nel primo caso si tratta di caratteri funzionali e nell'altro di caratteri strutturali.

Nella tavola XXXIX i dati si riferiscono a gruppi diversi di individui, ma, per ogni gruppo, sono considerati caratteri differenti (lunghezza, larghezza, circonferenza della testa). In questo caso, possiamo quindi scindere l'influenza dei due fattori.

Se calcoliamo la cograduazione tra i valori di detta tavola che lo stesso carattere presenta nei cinque differenti gruppi di individui, troviamo una cograduazione nulla tra la lunghezza e la circonferenza, ma una cograduazione positiva tra questa e la larghezza (cograduazione tra indice di ipernormalità e scostamento semplice medio = 0,33, tra indice di ipernormalità e scostamento quadratico medio = 0,67) e, ciò sia che si consideri lo scostamento percentuale o lo scostamento relativo.

Dei cinque indici di cograduazione, invece, che si ottengono per i diversi caratteri dei cinque gruppi di individui considerati, due (per i maschi di Teneriffa e di Gomera) risultano negativi (con valori di $-0,5$ e $-0,25$) e gli altri tre positivi (con valori di $0,5$, 1 , $0,5$). La media dei cinque indici è positiva ed uguale a $0,25$.

Sarebbe azzardato trarre conseguenze qualsiasi da questo solo esempio relativo a cinque gruppi e tre caratteri.

Esso può valere soltanto per indicare la via che si potrebbe seguire qualora le ricerche in argomento si volessero estendere ed approfondire.

Parecchi degli indici di ipernormalità che abbiamo calcolato nel corso di questo lavoro si prestano ad ulteriori osservazioni.

Nella tavola XL sono stati classificati gli indici di ipernormalità per le stature dei 25 cantoni svizzeri. Essi sono riprodotti singolarmente nella tavola seguente, accanto ai numeri delle osservazioni a cui si riferiscono e ai valori dello scostamento semplice medio 1S_A e dello scostamento quadratico medio 2S_A nonchè ai rispettivi rapporti di detti scostamenti alla media aritmetica ${}^1S_A/A$ e ${}^2S_A/A$.

Variabilità e ipernormalità delle distribuzioni dei coscritti dei cantoni svizzeri

CANTONE	NUMERO DEI CASI	$1S_A$	$2S_A$	$1S_{A/A}$	$2S_{A/A}$	I
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Zurigo	3.454	5,18	6,54	3,06	3,86	3,188
Berna.	8.723	5,18	6,67	3,07	3,96	3,316
Lucerna.	1.971	5,41	6,84	3,22	4,07	3,197
Uri.	274	5,11	6,56	3,06	3,94	3,296
Schwyz	713	5,18	6,70	3,08	3,99	3,346
Obwalden	206	5,15	6,76	3,07	4,03	3,446
Nidwalden.	219	5,01	6,27	2,96	3,71	3,132
Glarus	346	4,76	6,18	2,83	3,67	3,371
Zug.	237	5,41	6,72	3,20	3,97	3,086
Friburgo	1.467	5,41	7,02	3,23	4,19	3,368
Solothurn	1.139	5,13	6,66	3,03	3,94	3,371
Basilea città.	761	5,33	6,70	3,10	3,89	3,160
Basilea campagna	801	5,22	7,04	3,09	4,16	3,638
Sciaffusa	455	5,13	6,36	3,00	3,71	3,074
Appenzell A.-Rh.	670	5,40	6,84	3,25	4,12	3,209
Appenzell I.-Rh.	259	5,15	6,60	3,13	4,01	3,285
San Gallo.	2.627	5,08	6,62	3,04	3,96	3,396
Grigioni.	1.092	5,24	6,74	3,10	3,99	3,309
Aargau	2.971	5,22	6,84	3,09	44,05	3,434
Thurgau	1.320	5,29	6,86	3,14	4,07	3,363
Ticino	1.209	4,76	6,12	2,86	3,67	3,306
Waadt	1.997	5,10	6,46	3,01	3,81	3,209
Vallese	1.364	5,05	6,28	3,00	3,73	3,093
Neuenburg	698	5,09	6,38	3,01	3,77	3,142
Ginevra.	401	5,29	6,70	3,08	3,90	3,208
Svizzera.	35.374	5,25	6,74	3,12	4,01	3,285

Indice di cograduazione tra le colonne :

$$7 \text{ e } 3 = -0,098$$

$$7 \text{ e } 5 = -0,021$$

$$7 \text{ e } 4 = 0,232$$

$$7 \text{ e } 6 = 0,255$$

L'indice di cograduazione risulta positivo tra I e 2S_A e tra I e ${}^2S_A/A$ e negativo tra I e 1S_A e tra I e ${}^1S_A/A$ come si rileva dai valori in calce alla tavola. Ciò si può attribuire al fatto che, nella formula per il calcolo di I , 2S_A figura al numeratore e 1S_A al denominatore. Il fatto che la cograduazione risulti positiva col primo valore e negativa col secondo suggerisce che i due valori variano con una certa indipendenza, vale a dire che il rapporto $\frac{{}^2S_A}{{}^1S_A}$ risulta affetto da notevoli perturbazioni accidentali, evidentemente in relazione col numero limitato delle osservazioni in molti Cantoni.

In questo caso, si tratta dello stesso carattere (statura) in gruppi di popolazioni diverse dai vari cantoni.

Invece per le Faeröer abbiamo calcolato i valori di I per i vari caratteri della stessa popolazione. Essi sono indicati singolarmente nella tavola XLI, insieme ai valori degli indici percentuali $\frac{{}^1S_A}{A}$ e $\frac{{}^2S_A}{A}$ e degli indici relativi di variabilità $\frac{{}^1S_A}{\sqrt{A}}$ e $\frac{{}^2S_A}{\sqrt{A}}$. La cograduazione risulta negativa sia per i maschi che per le femmine con tutti gli indici di variabilità.

Nella tavola XXXI-*bis* sono stati esposti i valori di 1S_A , 2S_A , ed I per la statura, il peso e il perimetro toracico dei coscritti inglesi classificati secondo il grado di idoneità al servizio militare. Si tratta di caratteri strutturali, lineare il primo; non lineari gli altri due.

La cograduazione tra gli indici di ipernormalità e di variabilità risulta in ogni caso positiva, come si vede dagli indici in calce a detta tavola. La spiegazione di tale cograduazione positiva può essere trovata nel fatto che la ipernormalità delle distribuzioni si può ritenere determinata dalla presenza, a fianco del tipo normale, di due ectipi estremi tanto più importanti quanto minore è il grado di idoneità dei coscritti del gruppo considerato. Ora, questi ectipi evidentemente accrescono, oltre alla ipernormalità, la variabilità del carattere.

Infine, possiamo determinare l'indice di cograduazione per le stature degli studenti e delle studentesse dei Collegi americani (tavola XLII). In questo caso, gli indici di cograduazione per le femmine risultano una volta negativo e una volta positivo, mentre per i maschi essi sono sempre negativi.

TAVOLA XLI

*Variabilità e ipernormalità relative agli abitanti adulti delle isole
Faeröer*

CARATTERI	<i>A</i>	\sqrt{A}	1S_A	2S_A	$\frac{^1S_A}{A}$	$\frac{^1S_A}{\sqrt{A}}$	$\frac{^2S_A}{A}$	$\frac{^2S_A}{\sqrt{A}}$	<i>I</i>
1	2	3	4	5	6	7	8	9	10
diam. ant. post.	M 195,017	13,965	5,164	6,364	0,0265	0,3698	0,0326	0,4557	3,037
	F 185,242	13,647	5,063	6,281	0,0272	0,3710	0,0337	0,4602	3,079
diam. trasversale	M 186,206	13,646	4,532	5,778	0,0243	0,3321	0,0310	0,4234	3,251
	F 149,734	12,237	4,156	5,348	0,0278	0,3396	0,0357	0,4370	3,312
diam. front. minimo . . .	M 109,531	10,466	3,603	4,513	0,0329	0,3289	0,0412	0,4312	3,139
	F 104,860	10,240	3,307	4,103	0,0315	0,3229	0,0391	0,4007	3,080
diam. bizygomatico	M 145,846	12,077	4,187	5,160	0,0287	0,3467	0,0354	0,4272	3,037
	F 135,722	11,650	3,901	4,940	0,0287	0,3348	0,0364	0,4240	3,208
diam. ophryo - al- veolare	M 90,351	9,505	3,216	4,303	0,0356	0,3383	0,0476	0,4527	3,580
	F 85,727	9,259	3,194	3,996	0,0373	0,3449	0,0466	0,4315	3,131
diam. naso - men- toniero	M 121,096	11,004	4,943	6,216	0,0408	0,4492	0,0513	0,5649	3,163
	F 112,633	10,613	4,436	5,451	0,0394	0,4180	0,0484	0,5136	3,020
diam. capello - men- toniero	M 182,886	13,524	6,028	7,570	0,0330	0,4457	0,0414	0,5597	3,154
	F 171,947	13,113	6,242	7,711	0,0363	0,4760	0,0448	0,5880	3,052
diam. bimascellare	M 114,543	10,703	5,174	6,421	0,0452	0,4834	0,0561	0,5999	3,080
	F 104,107	10,203	3,853	4,923	0,0370	0,3776	0,0473	0,4825	3,262
lunghezza del naso	M 55,908	7,477	2,643	3,313	0,0473	0,3534	0,0593	0,4430	3,143
	F 51,222	7,157	2,574	3,233	0,0503	0,3596	0,0631	0,4517	3,156
larghezza del naso	M 37,590	6,131	2,282	2,855	0,0607	0,3722	0,0759	0,4656	3,131
	F 33,425	5,781	2,042	2,508	0,0611	0,3532	0,0750	0,4338	3,016
statura	M 168,816	12,993	4,489	5,656	0,0266	0,3455	0,0335	0,4353	3,177
	F 158,199	12,574	4,461	5,757	0,0282	0,3548	0,0364	0,4578	3,331

Indice di cograduazione tra i dati delle colonne :

Maschi

10 e 6 = — 0,067

10 e 8 = — 0,067

10 e 7 = — 0,333

10 e 9 = — 0,133

Femmine

10 e 6 = — 0,040

10 e 8 = — 0,333

10 e 6 = — 0,027

10 e 9 = — 0,167

Riassumendo, i risultati del confronto tra il comportamento dei caratteri funzionali e dei caratteri strutturali abbiamo constatato che essi differiscono sostanzialmente.

I caratteri funzionali presentano una ipernormalità più spiccata dei caratteri strutturali lineari.

Nei caratteri funzionali, l'indice di ipernormalità del carattere risulta superiore alla media degli indici dei suoi componenti; nei caratteri strutturali, invece, questo non si verifica.

TAVOLA XLII

Variabilità e ipernormalità delle distribuzioni delle stature degli studenti e delle studentesse dei Collegi americani

ETÀ	MASCHI			FEMMINE		
	1S_A	2S_A	I	1S_A	2S_A	I
	(1)	(2)	(3)	(4)	(5)	(6)
16 anni	2,301	2,888	3,149	1,806	2,266	3,148
17 "	2,128	2,709	3,242	1,820	2,291	3,170
18 "	2,079	2,642	3,223	1,865	2,345	3,161
19 "	2,038	2,604	3,265	1,857	2,356	3,218
20 "	2,131	2,705	3,222	1,851	2,349	3,223
21 "	2,107	2,698	3,279	1,867	2,333	3,123
da 16 a 21	2,104	2,677	3,238	1,851	2,333	3,177

Indice di cograduazione tra i dati delle colonne :

$$3 \text{ e } 1 = -0,556$$

$$6 \text{ e } 4 = -0,222$$

$$3 \text{ e } 2 = -0,444$$

$$6 \text{ e } 5 = 0,444$$

La cograduazione tra gli indici di ipernormalità e gli indici di variabilità è positiva per i caratteri funzionali; mentre, per i caratteri strutturali lineari, varia di caso in caso, e ciò, sia che si tratti dello stesso carattere in gruppi di popolazione diversi o di caratteri diversi per la stessa popolazione.

Per detti caratteri strutturali lineari, la cograduazione talvolta è positiva, altre volte è negativa ed avviene anche che sia positiva quando la variabilità è desunta dallo scostamento qua-

dratico medio e negativo quando essa è desunta dallo scostamento semplice medio.

Quest'ultimo caso si verifica quando, il numero dei casi in osservazione essendo ristretto, il rapporto tra lo scostamento quadratico medio e lo scostamento semplice medio risulta fortemente affetto da perturbazioni accidentali.

Una cograduazione positiva si verifica quando la più o meno alta ipernormalità dipende dalla maggiore o minore importanza che assumono, di fronte al tipo normale del carattere, gli ectipi estremi; i quali, insieme alla ipernormalità, rialzano anche la variabilità. Le cause del comportamento degli indici di cograduazione in altri casi non si sono potuti precisare.

Per ciò che concerne i caratteri strutturali non lineari (per es. circonferenza e volume), i risultati ottenuti suggeriscono che l'ipernormalità del carattere complessivo non sia superiore a quella dei caratteri componenti, come avviene nei caratteri strutturali lineari, ma che vi sia una correlazione positiva tra la variabilità e indice di ipernormalità. Si tratta di una semplice suggestione, perchè le serie esaminate sono troppo poche per autorizzare una conclusione.

PARTE TERZA

APPENDICI TEORICHE

(a cura della Dott.ssa ANITA TREZZA)

APPENDICE PRIMA

II. — *Iponormalità attribuita alla composizione di due distribuzioni normali con uguale variabilità e media diversa.*

I. — Siano Y_1 e Y_2 due distribuzioni normali con media $-\delta$ e $+\delta$ rispettivamente e con uguale variabilità, che possiamo supporre uguale a 1, se sommiamo queste due distribuzioni la funzione densità della distribuzione Y ottenuta è:

$$\frac{1}{2\sqrt{2\pi}} \left[e^{-\frac{1}{2}(x-\delta)^2} + e^{-\frac{1}{2}(x+\delta)^2} \right] dx \quad (I)$$

Vogliamo dimostrare che la distribuzione Y è ipornormale. Faremo vedere che l'indice del Gini I , uguale al doppio del quadrato del rapporto tra lo scarto quadratico medio e quello semplice, è per questa distribuzione minore di π e analogamente che l'indice del Pearson, uguale al rapporto tra il quarto momento e il quadrato del secondo momento intorno alla media aritmetica, è minore di 3. Esaminiamo l'indice di ipernormalità $I = \frac{2^2 S^2}{1 S^2}$ del Gini e dimostriamo che $I < \pi$.

Lo scarto semplice medio della distribuzione Y è

$$1S = \frac{2}{2\sqrt{2}\pi} \int_0^\infty x \left[e^{-\frac{1}{2}(x-\delta)^2} + e^{-\frac{1}{2}(x+\delta)^2} \right] dx \quad (2)$$

Se poniamo nella 2) $x - \delta = \xi$ e $x + \delta = \eta$ abbiamo:

$$\begin{aligned} 1S &= \frac{1}{\sqrt{2}\pi} \left[\int_{-\delta}^\infty (\xi + \delta) e^{-\frac{\xi^2}{2}} d\xi + \int_\delta^\infty (\eta - \delta) e^{-\frac{\eta^2}{2}} d\eta \right] = \quad (3) \\ &= \frac{1}{\sqrt{2}\pi} \left\{ \int_{-\delta}^\infty \xi e^{-\frac{\xi^2}{2}} d\xi + \delta \int_{-\delta}^\infty e^{-\frac{\xi^2}{2}} d\xi + \int_\delta^\infty \eta e^{-\frac{\eta^2}{2}} d\eta - \delta \int_\delta^\infty e^{-\frac{\eta^2}{2}} d\eta \right\} = \\ &= \frac{1}{\sqrt{2}\pi} \left\{ \int_{-\delta}^0 \xi e^{-\frac{\xi^2}{2}} d\xi + \int_0^\infty \xi e^{-\frac{\xi^2}{2}} d\xi + \int_0^\infty \eta e^{-\frac{\eta^2}{2}} d\eta - \int_0^\delta \eta e^{-\frac{\eta^2}{2}} d\eta + \right. \\ &\quad \left. + \delta \int_{-\delta}^0 e^{-\frac{\xi^2}{2}} d\xi + \delta \int_0^\infty e^{-\frac{\xi^2}{2}} d\xi - \delta \int_0^\infty e^{-\frac{\eta^2}{2}} d\eta + \delta \int_0^\delta e^{-\frac{\eta^2}{2}} d\eta \right\} = \\ &= \frac{1}{\sqrt{2}\pi} \left\{ 2 \int_0^\infty \xi e^{-\frac{\xi^2}{2}} d\xi - 2 \int_0^\delta \xi e^{-\frac{\xi^2}{2}} d\xi + 2 \int_0^\delta e^{-\frac{\eta^2}{2}} d\eta \right\} = \\ &= \frac{\sqrt{2}}{\sqrt{\pi}} \left\{ e^{-\frac{\delta^2}{2}} + \delta \int_0^\delta e^{-\frac{x^2}{2}} dx \right\} \end{aligned}$$

Il quadrato dello scarto quadratico medio della distribuzione Y è:

$$^2S^2 = \frac{2}{2\sqrt{2}\pi} \left[\int_0^{\infty} x^2 \left[e^{-\frac{1}{2}(x-\delta)^2} + e^{-\frac{1}{2}(x+\delta)^2} \right] dx \right] \quad (4)$$

Con la sostituzione $x - \delta = \xi$, $x + \delta = \eta$ la (4) diventa:

$$\begin{aligned} ^2S^2 &= \frac{1}{\sqrt{2}\pi} \left[\int_{-\delta}^{+\infty} \xi^2 e^{-\frac{\xi^2}{2}} d\xi + \delta^2 \int_{-\delta}^{\infty} e^{-\frac{\xi^2}{2}} d\xi + 2\delta \int_{-\delta}^{\infty} \xi e^{-\frac{\xi^2}{2}} d\xi + \right. \\ &\quad \left. + \int_{\delta}^{\infty} \eta^2 e^{-\frac{\eta^2}{2}} d\eta + \delta^2 \int_{\delta}^{\infty} e^{-\frac{\eta^2}{2}} d\eta - 2\delta \int_{\delta}^{\infty} \eta e^{-\frac{\eta^2}{2}} d\eta \right] = \\ &= \frac{1}{\sqrt{2}\pi} \left[\int_{-\infty}^{\delta} \xi^2 e^{-\frac{\xi^2}{2}} d\xi + \delta^2 \int_{-\infty}^{\delta} e^{-\frac{\xi^2}{2}} d\xi + \int_{\delta}^{\infty} \eta^2 e^{-\frac{\eta^2}{2}} d\eta + \delta^2 \int_{\delta}^{\infty} e^{-\frac{\eta^2}{2}} d\eta \right] \\ &= \frac{1}{\sqrt{2}\pi} \left[\int_{-\infty}^{\infty} \eta^2 e^{-\frac{\eta^2}{2}} d\eta + \delta^2 \int_{-\infty}^{+\infty} e^{-\frac{\xi^2}{2}} d\xi \right] = \\ &= \frac{1}{\sqrt{2}\pi} [\sqrt{2}\pi + \delta^2 \sqrt{2}\pi] = 1 + \delta^2 \end{aligned}$$

L'indice di ipernormalità della distribuzione Y calcolato come doppio del quadrato del rapporto tra lo scarto quadratico medio e quello semplice medio è:

$$I = \frac{^2S^2}{^1S^2} = \pi \frac{1 + \delta^2}{\left(e^{-\frac{\delta^2}{2}} + \delta \int_0^{\delta} e^{-\frac{x^2}{2}} dx \right)^2} \quad (5)$$

Vogliamo dimostrare che $I < \pi$

Per questo basta dimostrare che ;

$$\frac{1 + \delta^2}{\left(e^{-\frac{\delta^2}{2}} + \delta \int_0^{\delta} e^{-\frac{x^2}{2}} dx \right)^2} < 1 \quad (6)$$

cioè :

$$\left(e^{-\frac{\delta^2}{2}} + \delta \int_0^{\delta} e^{-\frac{x^2}{2}} dx \right)^2 > 1 + \delta^2 \quad (7)$$

2. — Premettiamo alcune considerazioni che ci saranno utili per la dimostrazione della disuguaglianza (7)

Dimostriamo che se $x > 0$:

$$\sum_{k=0}^{2p+1} \frac{(-1)^k}{k!} \left(\frac{x^2}{2} \right)^k < e^{-\frac{x^2}{2}} < \sum_{k=0}^{2p} \frac{(-1)^k}{k!} \left(\frac{x^2}{2} \right)^k \quad (8)$$

E infatti $e^{-\frac{x^2}{2}} < 1$, quindi la funzione $f(x) = e^{-\frac{x^2}{2}} - 1 + \frac{x^2}{2}$, nulla per $x = 0$, è positiva per $x > 0$, essendo la sua derivata $f'(x) = x(1 - e^{-\frac{x^2}{2}})$ positiva per $x > 0$.

Si ha quindi :

$$1 - \frac{x^2}{2} < e^{-\frac{x^2}{2}} < 1 \quad (9)$$

Ammettiamo di aver stabilito che :

$$e^{-\frac{x^2}{2}} < \sum_{k=0}^{2p} \frac{(-1)^k}{k!} \left(\frac{x^2}{2} \right)^k$$

e consideriamo la funzione

$$e^{-\frac{x^2}{2}} - \sum_{k=0}^{2p+1} \frac{(-1)^k}{k!} \left(\frac{x^2}{2} \right)^k \quad (10)$$

essa è nulla per $x = 0$ ed è positiva per $x > 0$, infatti la sua derivata uguale a

$$x \left[-e^{-\frac{x^2}{2}} + \sum_{k=0}^{2p} \frac{(-1)^k}{(k)!} \left(\frac{x^2}{2} \right)^k \right]$$

è senz'altro positiva per l'ipotesi iniziale.

Quindi

$$e^{-\frac{x^2}{2}} > \sum_{k=0}^{2p+1} \frac{(-1)^k}{k!} \left(\frac{x^2}{2} \right)^k$$

Se avessimo supposto

$$e^{-\frac{x^2}{2}} > \sum_{k=0}^{2p} \frac{(-1)^k}{k!} \left(\frac{x^2}{2} \right)^k$$

saremmo giunti alla conseguenza

$$e^{-\frac{x^2}{2}} < \sum_{k=0}^{2p+1} \frac{(-1)^k}{k!} \left(\frac{x^2}{2} \right)^k$$

Questo è evidentemente assurdo perchè :

$$\sum_{k=0}^{2p} \frac{(-1)^k}{k!} \left(\frac{x^2}{2} \right)^k > \sum_{k=0}^{2p+1} \frac{(-1)^k}{k!} \left(\frac{x^2}{2} \right)^k$$

3 — Ci proponiamo ora di dimostrare la disuguaglianza (7) del paragrafo I. Notiamo intanto che posto :

$$F(\delta) = \left(e^{-\frac{\delta^2}{2}} + \delta \int_0^\delta e^{-\frac{x^2}{2}} dx \right)^2 - 1 - \delta^2 \quad (11)$$

si ha che $F(0) = 0$ e quindi se dimostriamo che per qualsiasi valore di $\delta > 0$ si ha :

$$\frac{dF(\delta)}{d\delta} > 0 \quad (12)$$

è anche

$$F(\delta) > 0.$$

Per la dimostrazione fatta al paragrafo precedente si ha che :

$$e^{-\frac{\delta^2}{2}} > 1 - \frac{\delta^2}{2} + \frac{1}{2!} \left(\frac{\delta^2}{2} \right)^2 - \frac{1}{3!} \left(\frac{\delta^2}{2} \right)^3$$

e per conseguenza del fatto che

$$e^{-\frac{x^2}{2}} > 1 - \frac{x^2}{2}$$

si ha :

$$\int_0^\delta e^{-\frac{x^2}{2}} dx > \int_0^\delta \left(1 - \frac{x^2}{2} \right) dx = \delta - \frac{\delta^3}{6}$$

Si ha che :

$$\frac{dF(\delta)}{d\delta} = 2e^{-\frac{\delta^2}{2}} \int_0^\delta e^{-\frac{x^2}{2}} dx + 2\delta \left[\int_0^\delta e^{-\frac{x^2}{2}} dx \right]^2 - 2\delta \quad (13)$$

Quindi per quello che abbiamo sopra detto si ha

$$\begin{aligned} \frac{dF(\delta)}{d\delta} &= 2 \left[\int_0^\delta e^{-\frac{x^2}{2}} dx \left(e^{-\frac{\delta^2}{2}} + \delta \int_0^\delta e^{-\frac{x^2}{2}} dx \right) - \delta \right] > \\ &> 2 \left(\delta - \frac{\delta^3}{6} \right) \left(1 + \frac{\delta^2}{2} - \frac{\delta^4}{24} - \frac{\delta^6}{48} \right) - 2\delta = \\ &= \frac{1}{144} (-\delta^6 - 2\delta^4 + 24\delta^2 + 48) (6\delta - \delta^3) - 2\delta = \\ &= \frac{1}{144} (\delta^9 - 4\delta^7 - 36\delta^5 + 96\delta^3) = \frac{\delta^3}{144} (\delta^6 - 4\delta^4 - 36\delta^2 + 96) \end{aligned} \quad (14)$$

Dalla 13) si può subito vedere che se

$$\left[\int_0^\delta e^{-\frac{x^2}{2}} dx \right]^2 > 1$$

la disuguaglianza (12) è senz'altro verificata.

Per $\delta = 1,3$ si ha

$$\left[\int_0^{\delta} e^{-\frac{x^2}{2}} dx \right]^2 = 1,17$$

quindi basterà dimostrare la 12 per

$$0 < \delta < 1,3.$$

Ma per

$$0 < \delta < 1,3$$

si ha senz'altro

$$\frac{dF(\delta)}{d\delta} > \frac{\delta^3}{144} (\delta^6 - 4\delta^4 - 36\delta^2 + 96) > 0$$

Quindi è dimostrato che per $\delta > 0$ si ha

$$\frac{dF(\delta)}{d\delta} > 0$$

e di conseguenza

$$F(\delta) > 0$$

Abbiamo così dimostrato che la distribuzione ottenuta sommando due distribuzioni normali con media diversa e uguale variabilità è iponormale.

L'indice di ipernormalità del Gini è risultato infatti per questa distribuzione minore di π .

4. Consideriamo ora come indice di ipernormalità quello del Pearson cioè $\frac{\mu_4}{(\mu_2)^2}$, dove μ_4 indica il momento quarto e μ_2 il

momento secondo e dimostriamo che una distribuzione ottenuta componendo due distribuzioni normali con media diversa e uguale variabilità ha l'indice del Pearson minore di tre.

Con le stesse notazioni usate nel paragrafo precedente la funzione densità di questa nuova distribuzione è:

$$\frac{1}{2\sqrt{2}\pi} \left[e^{-\frac{1}{2}(x-\delta)^2} + e^{-\frac{1}{2}(x+\delta)^2} \right] dx \quad (1)$$

Il momento quarto di questa distribuzione è :

$$\mu'_4 = \frac{1}{2\sqrt{2}\pi} \int_0^{\infty} x^4 \left[e^{-\frac{1}{2}(x-\delta)^2} + e^{-\frac{1}{2}(x+\delta)^2} \right] dx \quad (2)$$

Facciamo la sostituzione

$$x - \delta = \xi, \quad x + \delta = \eta$$

$$2\sqrt{2}\pi\mu'_4 = \int_{-\delta}^{\infty} (\xi + \delta)^4 e^{-\frac{1}{2}\xi^2} d\xi + \int_{+\delta}^{\infty} (\eta - \delta)^4 e^{-\frac{1}{2}\eta^2} d\eta \quad (3)$$

A calcoli effettuati, analogamente a ciò che si è fatto nel paragrafo I, e posto uguale a μ_4 il momento quarto delle due distribuzioni Y_1 e Y_2 e uguale a μ_2 il loro momento secondo, che è 1 per ipotesi si ha :

$$\mu'_4 = \mu_4 + \delta^4 + 6\delta^2\mu_2$$

Ed essendo $\mu_2 = 1$ si ha che $\mu_4 = 3$, essendo le due distribuzioni normali e quindi si ha

$$\mu'_4 = 3 + \delta^4 + 6\delta^2 \quad (4)$$

Il momento secondo della distribuzione $Y = Y_1 + Y_2$ è come abbiamo già visto nel paragrafo precedente :

$$\mu'_2 = 1 + \delta^2 \quad (5)$$

L'indice del Pearson della distribuzione Y è

$$\frac{\mu_4}{\mu_2^2} = \frac{3 + \delta^4 + 6\delta^2}{(1 + \delta^2)^2} \quad (6)$$

Dimostriamo che è

$$\frac{\mu_4}{\mu_2^2} < 3 \quad (7)$$

Infatti è vera la disuguaglianza seguente :

$$3 + \delta^4 + 6 \delta^2 < 3 + 6 \delta^2 + 3 \delta^4 \quad (8)$$

che è equivalente alla (7)

APPENDICE SECONDA

12. — *Ipernormalità attribuita alla composizione di due distribuzioni aventi uguale media e diversa variabilità.*

1. — Dimostriamo che la distribuzione ottenuta sommando due distribuzioni Y_1 e Y_2 aventi uguale media e diversa variabilità è ipernormale. Dimosteremo che l'indice di ipernormalità adottato dal Gini $I = 2 \frac{S^2}{S^2}$ per questa distribuzione è maggiore di π e nel paragrafo 5) faremo vedere che l'indice del Pearson è maggiore di 3.

Possiamo far coincidere la media delle distribuzioni Y_1 e Y_2 con l'origine delle coordinate e considerare uguale ad 1 la variabilità della prima distribuzione ed uguale a k quella della seconda, essendo $k > 1$.

Le due distribuzioni sono quindi :

$$Y_1 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}x^2} dx$$

$$Y_2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{x}{k}\right)^2} dx$$

La media della nuova distribuzione è l'origine delle coordinate e la funzione densità è :

$$\frac{1}{2\sqrt{2\pi}} \left[e^{-\frac{1}{2}x^2} + e^{-\frac{1}{2}\frac{x^2}{k^2}} \right]$$

Il quadrato dello scostamento quadratico medio della nuova distribuzione è :

$$^2S^2 = \frac{1}{2\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 \left[e^{-\frac{1}{2}x^2} + e^{-\frac{1}{2}\frac{x^2}{k^2}} \right] dx = \frac{1}{2} (1 + k^2)$$

Se indichiamo con 1S_1 e 1S_2 gli scostamenti semplici medi di Y_1 e Y_2 , lo scostamento semplice medio della nuova distribuzione è

$$^1S = \frac{1}{2\sqrt{2\pi}} \int_{-\infty}^{+\infty} x \left[e^{-\frac{x^2}{2}} + e^{-\frac{1}{2}\frac{x^2}{k^2}} \right] dx = \frac{^1S_1 + ^1S_2}{2} = \frac{1}{2} \sqrt{\frac{2}{\pi}} (1 + k)$$

Infatti, essendo Y_1 e Y_2 due distribuzioni normali, si ha che :

$$^1S_1 = \sqrt{\frac{2}{\pi}} \quad ^2S_1 = \sqrt{\frac{2}{\pi}} \quad \text{e} \quad ^1S_2 = \sqrt{\frac{2}{\pi}} \quad ^2S_2 = \sqrt{\frac{2}{\pi}} k$$

dove abbiamo indicato con 2S_1 e 2S_2 gli scostamenti quadratici medi delle due distribuzioni.

L'indice d'ipernormalità è quindi :

$$I = \frac{2 \ ^2S^2}{^1S^2} = \pi \frac{2 (1 + k^2)}{1 + k^2 + 2k}$$

Si dimostra immediatamente che :

$$\frac{2 + 2k^2}{1 + k^2 + 2k} > 1 \quad (1)$$

Infatti questa disuguaglianza equivale alla seguente :

$$2 + 2k^2 > 1 + k^2 + 2k \quad (2)$$

cioè :

$$k^2 - 2k + 1 = 0 \quad (3)$$

La 3) è senz'altro verificata, essendo il primo membro uguale a $(k-1)^2$.

2. — Vogliamo vedere che influenza ha sull'indice di ipernormalità della distribuzione la differenza tra le variabilità delle due distribuzioni Y_1 e Y_2 .

Esaminiamo come varia al crescere di $k > 1$ l'indice di ipernormalità $I = 2 \frac{{}^2S^2}{{}^1S^2}$ per questo esaminiamo la seguente funzione:

$$\frac{dI}{dk} = 2\pi \frac{d}{dk} \left(\frac{1 + k^2}{1 + k^2 + 2k} \right) = 2\pi \frac{2k^2 - 2}{(1 + k^2 + 2k)^2} \quad (1)$$

La funzione $\frac{dI}{dk}$ è maggiore di zero poichè noi abbiamo supposto che $k > 1$. L'indice di ipernormalità cresce quindi al crescere di k e tende a 2π , per k che tende ad un valore infinitamente grande.

Per $k = 1$, come è evidente, si ha il minimo valore dell'indice e precisamente π . La somma di due distribuzioni normali uguali è infatti una distribuzione normale.

3. — Consideriamo la distribuzione $Y = Y_1 + n Y_2$ essendo $n > 0$, ed esaminiamo come varia l'indice di ipernormalità al variare di n . La funzione densità della distribuzione Y è:

$$\frac{1}{(n+1)\sqrt{2\pi}} \left[e^{-\frac{1}{2}x^2} + n e^{-\frac{1}{2} \frac{x^2}{k^2}} \right]$$

Lo scostamento semplice medio è:

$$\begin{aligned} {}^1S &= \frac{1}{(n+1)\sqrt{2\pi}} \int_{-\infty}^{+\infty} \left[e^{-\frac{1}{2}x^2} + n e^{-\frac{x^2}{2k^2}} \right] dx = \\ &= \frac{1}{(n+1)} ({}^1S_1 + n {}^1S_2) = \frac{1}{n+1} \sqrt{\frac{2}{\pi}} (1 + nk) \end{aligned} \quad (1)$$

Il quadrato dello scostamento quadratico medio è:

$${}^2S^2 = \frac{1}{(n+1)\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 \left[e^{-\frac{1}{2}x^2} + n e^{-\frac{x^2}{2k^2}} \right] dx = \frac{1}{n+1} (1 + nk^2) \quad (2)$$

L'indice di ipernormalità della distribuzione Y è quindi :

$$I_n = \frac{2^3 S^2}{1 S_2} = \pi \frac{(n+1)(1+nk^2)}{1+n^2 k^2+2nk} \quad (3)$$

È chiaro che I_n è maggiore di π . Infatti la diseguaglianza

$$(n+1) \frac{1+nk^2}{1+n^2 k^2+2nk} > 1$$

è equivalente alla seguente :

$$n+1+n^2 k^2+nk^2 > 1+n^2 k^2+2nk$$

cioè

$$nk^2+n+2nk-0$$

che è vera essendo

$$nk^2+n+2nk=n(k+1)^2 > 0$$

Vediamo che influenza ha sull'indice di ipernormalità il numero n .

Esaminiamo per quali valori di n I_n è funzione crescente di n . Notiamo che :

$$I_n = \pi \left[1 + \frac{n(1-k)^2}{1+n^2 k^2+2nk} \right] \quad (4)$$

Per essere I_n funzione crescente di n deve essere :

$$\frac{d}{dn} \left(\frac{n}{1+n^2 k^2+2nk} \right) > 0$$

Esaminiamo per quali valori di n questo accade.

$$\frac{d}{dn} \left(\frac{n}{1+n^2 k^2+2nk} \right) = \frac{1-n^2 k^2}{(1+nk)^4}$$

Si ha che :

$$1-n^2 k^2 > 0$$

Se :

$$n k < 1$$

cioè se :

$$n < \frac{1}{k}$$

I_n è funzione crescente di n . Ricordiamo che $k > 1$ quindi se $n > 1$ l'indice di ipernormalità decresce sempre e :

$$\lim_{n \rightarrow \infty} I_n = \lim_{n \rightarrow \infty} \pi \left(1 + \frac{n (1 - k)^2}{1 + n^2 k^2 + 2 n k} \right) = \pi$$

4. — Consideriamo ora la distribuzione

$$Y = n Y_1 + Y_2 \quad (1)$$

Con semplici calcoli si trova che :

$$I_n = \frac{\pi (n + 1) (n + k^2)}{n^2 + k^2 + 2 n k} = \pi \left[1 + \frac{n (k - 1)^2}{n^2 + k^2 + 2 n k} \right] \quad (2)$$

Per essere I_n funzione crescente di n deve essere :

$$\frac{d}{dn} \left(\frac{n}{n^2 + k^2 + 2 n k} \right) > 0 \quad (3)$$

Esaminiamo per quali valori di n questo accade.

$$\frac{d}{dn} \left(\frac{n}{n^2 + k^2 + 2 n k} \right) = \frac{k^2 - n^2}{(n + k)^4} \quad (4)$$

Se $n < k$, I_n è funzione crescente di n , se $n > k$, I_n è funzione decrescente di n . Il massimo valore che può assumere I_n al variare di n , è precisamente quello che assume per $n = k$ e vale :

$$I_{max} = \pi \left[1 + \frac{k (k - 1)^2}{4 k^2} \right] = \pi \frac{(k + 1)^2}{4 k} \quad (5)$$

Dato che la distribuzione Y considerata sopra si può scrivere :

$$Y = n \left(y_1 + \frac{1}{n} y_2 \right) = n (y_1 + n' y_2) \quad (6)$$

possiamo sostituire nella 2) del paragrafo precedente a n , $n' = \frac{1}{n}$

Si trova così :

$$I_{n'} = \pi \left[1 + \frac{n'^2 (1 - k^2)^2}{n'^2 + k^2 + 2 n' k} \right] \quad (7)$$

Si è trovata quindi per $I_{n'}$ la stessa espressione (2)

5. — Dimostriamo ora che l'indice del Pearson della distribuzione ottenuta sommando due distribuzioni aventi uguale media e diversa variabilità è maggiore di tre. La funzione densità di questa distribuzione è, come abbiamo già visto nel paragrafo 1)

$$\frac{1}{2\sqrt{2}\pi} \left[e^{-\frac{1}{2}x^2} + e^{-\frac{1}{2}\frac{x^2}{k^2}} \right]$$

Il momento quarto μ'_4 è :

$$\mu'_4 = \frac{1}{2\sqrt{2}\pi} \int_{-\infty}^{+\infty} x^4 \left[e^{-\frac{1}{2}x^2} + e^{-\frac{1}{2}\frac{x^2}{k^2}} \right] dx = \frac{1}{2} (3 + 3k^4)$$

L'indice del Pearson è quindi uguale al seguente rapporto :

$$\frac{\mu_4}{(\mu_2)^2} = \frac{2(3 + 3k^4)}{(1 + k^2)^2}$$

Dimostriamo che il rapporto $\frac{\mu_4}{(\mu_2)^2}$ è maggiore di 3, cioè che :

$$3(2 + 2k^4) > 3(1 + k^4 + 2k^2)$$

Questa disegualianza equivale alla seguente :

$$2 + 2k^4 > 1 + k^4 + 2k^2$$

da cui deriva:

$$1 + k^4 - 2k^2 > 0$$

diseguaglianza senz'altro vera.

Abbiamo così dimostrato che la distribuzione ottenuta sommando due distribuzioni aventi uguale media e diversa variabilità è ipernormale.

Calcolati infatti gli indici di ipernormalità del Gini e del Pearson, essi sono risultati maggiori del valore normale.

RIASSUNTO

Questo articolo condensa i risultati di minuziose ricerche sulla forma della distribuzione dei caratteri antropologici pazientemente continuate per quasi un trentennio presso la Facoltà di Scienze Statistiche, demografiche ed attuariali della Università di Roma e il Comitato italiano per lo studio dei problemi della Popolazione.

Nell'intervallo tra le due guerre mondiali, parve argomento adatto per le esercitazioni, che erano obbligatorie per gli studenti iscritti al corso di Statistica elementare nella Facoltà di scienze statistiche, demografiche ed attuariali della Università di Roma, l'interpolazione delle distribuzioni, delle stature dei coscritti presentatisi alle leve di terra, per i quali le relazioni annuali del Ministero della Guerra fornivano una lunga serie (dalla leva del 1854 a quella del 1920) di dati dettagliati tra loro comparabili.

Si trattava di interpolare, sulla distribuzione delle stature dei coscritti di ogni circoscrizione (compartimento o provincia o circondario) una curva normale, misurando poi l'accostamento che i dati teorici presentavano rispetto ai dati effettivi e la curtosi o ipernormalità della distribuzione.

Ci si è presto resi conto che, in contrasto con l'opinione corrente che riguardava la statura come esempio tipico di un carattere che segue nella sua distribuzione la curva normale, questa invece risultava quasi sempre ipernormale, le frequenze effettive essendo più frequenti delle teoriche al centro, intorno alla media, e ai due estremi della distribuzione.

La misura dell'ipernormalità fu basata sul rapporto tra lo scostamento quadratico medio e lo scostamento semplice medio dalla media aritmetica, in quanto queste due costanti sono di facile calcolo e facile riesce pure il controllo delle operazioni necessarie per ottenerle, controllo che è indispensabile per le esercitazioni affidate agli studenti. L'indice di iper-

normalità è precisamente desunto dalla quantità $I = 2 \left(\frac{2S}{1S} \right)^2$ che risulta ≈ 3.1416 a seconda che la distribuzione è ipernormale, normale o iponormale.

Per decidere se l'ipernormalità della distribuzione delle stature costituiva una peculiarità delle statistiche italiane o aveva invece valore generale, il calcolo degli indici di ipernormalità fu esteso ai dati delle statistiche dei coscritti francesi, analoghe a quelle dell'Italia, e a dati, che in quegli anni erano venuti via via pubblicandosi, relativi alla Svizzera ed all'Austria (distretto di Mistelbach). I valori degli indici risultavano di circa 3,20-3,30.

Ne risultò confermata la ipernormalità della distribuzione delle stature dei coscritti, che assumeva così il valore di una regolarità statistica. Questa è stata segnalata in una comunicazione presentata alla Società italiana di Statistica nella seduta del giugno 1949 e pubblicata, con qualche riduzione, in inglese l'anno successivo negli «Acta genetica et statistica medica» di Uppsala.

Continuavano frattanto le elaborazioni di nuovi dati relativi a coscritti.

Per l'Italia, veniva portato a 166 il numero delle distribuzioni delle stature relative a tutto il Paese o a singoli compartimenti. Tutte, salvo una, risultavano ipernormali.

Per la Svizzera, venivano calcolati gli indici per i singoli Cantoni, di cui, su 25, 21 risultavano ipernormali.

Per la Svizzera e per l'Italia, si è potuto constatare che i valori degli indici relativi a varie circoscrizioni territoriali non differiscono sistematicamente dal valore dell'indice relativo al territorio complessivo. Per la Francia e per l'Italia, si è analogamente constatato che gli indici ottenuti per i singoli anni non differiscono dall'indice per l'intero periodo.

Per il Belgio, vennero calcolati gli indici relativi ai coscritti del 1931, non solo per la distribuzione delle stature, ma anche dei perimetri toracici e dei pesi, ottenendo, per questi due ultimi caratteri, valori (3,41 e rispettivamente 3,36) sensibilmente superiori a quello relativo alla statura (3,17).

Per il Belgio, potevano anche essere considerati a parte i coscritti dichiarati abili, i quali mostravano per tutti e tre i caratteri, indici di ipernormalità, notevolmente inferiori a quelli per il complesso dei coscritti esaminati. È, questo, l'effetto della selezione, che elimina i valori estremi, soprattutto i più bassi, ma anche, in una certa misura, i più alti.

Per ciò che riguarda la statura, l'effetto della selezione venne messo in luce anche per parecchi altri paesi: Italia, Scozia, Finlandia, Svezia, Stati Uniti d'America, calcolando gli indici di ipernormalità per le distribuzioni delle stature dei militari in servizio, le quali risultano spesso iponormali.

Una eccezione è rappresentata dalla distribuzione delle stature dei soldati finlandesi alla fine del secolo XVIII e al principio del XIX che presenta indici di ipernormalità elevata, ma essa può essere attribuita, almeno in parte, al fatto che a quel tempo, come le fonti dichiarano, le visite mediche non erano rigorose.

Tutte queste conclusioni si basano su distribuzioni relative a coetanei (coscritti, reclute, soldati).

Diversi sono i risultati per distribuzioni che comprendono tutta la popolazione adulta.

Si sono potuti raccogliere dati per la popolazione adulta complessiva, e separatamente per i maschi e per le femmine, relativamente ai villaggi della Svizzera: Engtlingen e Frutigtal (Oberland), Untertoggenburg (San Gallo), Schangau (Oberemmenthal) e della Germania (Genkingen, Bregenzer Wald e paesi vicini), nonchè per Lapponi, Eschimesi, abitanti delle isole Faeröer e di alcune popolazioni della Melanesia (isole di Lir, quattro villaggi delle coste settentrionali della Nuova Guinea e quarantadue località della Nuova Irlanda) e della Micronesia (Kapingamarangi).

Molto spesso gli indici per le popolazioni complessive risultano iponormali, ma le loro medie in generale risultano ipernormali per il fatto che i singoli valori ipernormali si scostano dal valore medio per scostamenti più forti di quelli dei singoli indici iponormali.

Le ricerche relative alla statura e, per alcuni paesi, al peso e al perimetro toracico hanno potuto essere estese a molteplici altri caratteri per i coscritti della Svizzera e per gli adulti dei suddetti villaggi della stessa Svizzera, della Germania e delle isole Faeröer, della Melanesia e della Micronesia. Salvo per le Faeröer, le rilevazioni di tutti questi dati sono state eseguite dallo Schlaginhaufen, così che esse sono rigorosamente comparabili, e riguardano molteplici caratteri quantitativi (34 per i coscritti svizzeri e sempre parecchie decine per gli adulti dei vari villaggi), in modo che alle conclusioni che se ne traggono non è azzardato attribuire valore generale per i caratteri antropometrici.

Per i coscritti, gli indici risultano in grande maggioranza ipernormali, mentre per le popolazioni adulte essi sono molto spesso iponormali.

Gli indici ottenuti per 17 caratteri di coscritti della Svizzera si sono potuti mettere a confronto, carattere per carattere, con quelli degli adulti dei due sessi dei villaggi di questo e degli altri paesi. Una differenza risulta, nella gran maggioranza dei casi, evidente, in quanto che, 28 volte su 34, l'ipernormalità risulta superiore per le distribuzioni relative ai coscritti. Questo risultato appare confermato dai calcoli eseguiti sui dati forniti da altri autori per gli adulti delle isole Faeröer, degli Eschimesi, dei Samaritani e delle antiche popolazioni delle isole Canarie per quanto riguarda i caratteri cranici.

La differenza tra coetanei e popolazione adulta in generale si può spiegare col fatto che questa contiene individui di varie età e con l'età, pure fra gli adulti, variano i valori medi di molti caratteri.

Per alcune popolazioni (Italiani, Lapponi, studenti dei collegi americani, abitanti di qualcuna delle isole Canarie), si sono anche potuti calcolare gli indici di ipernormalità per diversi gruppi di età.

Nessuna differenza sistematica è dato cogliere tra gli indici relativi ai maschi e alle femmine, che si sono potuti calcolare per varie popolazioni (villaggi svizzeri, villaggi tedeschi, della Melanesia e della Micronesia, Turchi,

Samaritani, Lapponi, studenti americani, abitanti delle isole Faeröer e di alcune isole Canarie).

Si è anche cercato di stabilire se vi è una relazione tra il grado di ipernormalità delle distribuzioni e la dipendenza del carattere in misura maggiore o minore dall'eredità o dall'ambiente (naturale o sociale), giungendo alla conclusione che l'ipernormalità risulta in generale maggiore per i caratteri che maggiormente dipendono dall'ambiente.

Infine, si è potuto mettere in luce, su dati relativi alle reclute inglesi che, per gli individui difettosi, la variabilità (desunta dallo scostamento semplice medio e dallo scostamento quadratico medio) e la ipernormalità delle distribuzioni delle stature, dei pesi e dei perimetri toracici risulta più elevata che per gli individui normali, tanto più elevata quanto più il difetto è grave, risultato che è in armonia con la teoria del Quetelet e dei costituzionalisti italiani secondo cui l'uomo medio è anche il tipo dell'uomo sano.

Una trattazione a parte venne fatta per le distribuzioni della rifrazione dell'occhio, che mostrano un'ipernormalità assolutamente eccezionale in confronto a tutti gli altri caratteri finora studiati.

Furono a tale scopo calcolati gli indici di ipernormalità per distribuzioni relative a pazienti di cliniche oculistiche o anche a personale medico o militare, esaminati a Tubinga (Germania), Chicago (USA), Goteborg (Svezia), Uppsala (Svezia), Londra (Inghilterra), nonché a Negri del Gabon (Africa) e agli Eschimesi di Anmagssalik (Groenlandia).

I dati non sono tra loro rigorosamente comparabili, sia perchè nella maggior parte delle rilevazioni, sono state eliminate alcune categorie di occhi e, d'altra parte, diversa è la composizione degli individui esaminati secondo la occupazione, sia perchè i metodi di rilevazione non sono sempre stati gli stessi. Comunque, tutte le distribuzioni mostrano una spiccata ipernormalità, generalmente molto superiore a quella degli altri caratteri antropometrici.

Il risultato appare a tutta prima singolare, tanto più che tutte le misure dell'occhio da cui la rifrazione risulta (curvatura della cornea, curvatura della superficie esterna ed interna del cristallino e lunghezza dell'asse dell'occhio, quest'ultima alla sua volta dipendente dalla profondità della camera anteriore, dallo spessore del cristallino e dalla profondità della camera posteriore) apparivano avere una distribuzione normale. Ricerche più accurate mostrarono però che in realtà, almeno per ciò che concerne la lunghezza dell'asse, questo risultato dipendeva dal fatto che venivano trascurati i casi estremi. Quando essi vennero debitamente considerati, la distribuzione della lunghezza dell'asse risultò nettamente ipernormale, sempre meno, però, della rifrazione dell'occhio.

Una rappresentazione grafica della distribuzione della rifrazione e della lunghezza dell'asse su carta di probabilità ha messo in luce come esse si possano plausibilmente considerare come risultanti dalla sovrapposizione di tre curve componenti, una relativa al tipo emетропо ossia normale, e normale pure (cioè gaussiana) nella forma, le altre due corrispondenti ai due ectipi della miopia e della ipermetropia.

La maggiore ipernormalità della rifrazione dell'occhio, in confronto a quella della lunghezza dell'asse, si spiega con il fatto che, fin tanto che le variazioni della lunghezza dell'asse si mantengono entro i limiti normali, una compensazione si verifica tra esse e le variazioni degli altri fattori della rifrazione dell'occhio (curvatura della cornea e della superficie esterna ed interna del cristallino), mentre, quando la lunghezza dell'asse raggiunge dimensioni superiori o inferiori ad un dato limite, tale compensazione non si verifica. Ne viene che, nella distribuzione relativa alla rifrazione dell'occhio, mentre vengono conservati i due ectipi dipendenti dalle lunghezze ectipiche della lunghezza dell'asse, si verifica, nei valori normali della rifrazione, un forte addensamento intorno al valore medio. Di qui i valori eccezionalmente elevati dell'indice di ipernormalità.

A questo punto, ci si è domandato se tale spiegazione possa essere estesa anche all'ipernormalità che, sia pure in misura minore, si verifica per gli altri caratteri antropometrici.

Una distinzione è apparsa necessaria, a questo proposito, tra caratteri funzionali e caratteri strutturali.

I caratteri funzionali derivano dall'azione combinata di vari fattori che possono compensarsi nei loro effetti, così che, per essi, la spiegazione adottata per la rifrazione dell'occhio appare pure possibile. Questa conclusione si può riguardare come confermata da qualche riscontro finora eseguito su caratteri funzionali, che ha mostrato come l'ipernormalità del carattere risulti superiore a quella dei fattori componenti.

I caratteri strutturali lineari, invece, risultano costituiti da componenti che si sommano, ed è noto che i segmenti di uno stesso organo non si compensano nè risultano indipendenti nelle loro variazioni, ma, in una certa misura, sono legati da vincoli di solidarietà, che si esprimono in coefficienti di correlazione positivi, ma sempre notevolmente inferiori all'unità.

Si spiega così che, a differenza di quanto avviene fra i caratteri funzionali, i caratteri strutturali lineari risultanti non presentino una ipernormalità superiore a quella dei fattori componenti.

I caratteri funzionali e i caratteri strutturali lineari si comportano diversamente anche sotto altri aspetti.

L'ipernormalità è notevolmente più alta per i caratteri funzionali.

La cograduazione tra ipernormalità e variabilità è positiva per i caratteri funzionali, mentre, per i caratteri strutturali lineari, talvolta è pure positiva, ma altre volte negativa ed in qualche caso risulta negativa se la variabilità si misura mediante lo spostamento semplice medio e invece positiva se si misura mediante lo spostamento quadratico medio. Qualche volta è possibile spiegare il particolare comportamento della cograduazione nel caso dei caratteri strutturali lineari, altre volte no.

Per ciò che concerne i caratteri strutturali non lineari (circonferenza, volume ecc.), essi, nei casi che si sono potuti esaminare, sotto alcuni aspetti si comportano come i caratteri funzionali, sotto altri come i caratteri strutturali lineari.

La Dr. Anita Trezza, nelle Appendici, mostra che, addizionando due distribuzioni normali contenenti il medesimo numero di casi, si ottiene

una distribuzione iponormale se le due distribuzioni date hanno la stessa variabilità e medie differenti e, al contrario, una distribuzione ipernormale se le due distribuzioni date hanno la stessa media e differenti variabilità, e ciò sia che si adotti quale misura della ipernormalità l'indice del Pearson oppure quello del Gini. Per il secondo caso, essa considera anche l'ipotesi che il numero dei casi contenuti in una distribuzione sia differente da quello contenuto nell'altra e studia l'influenza che il rapporto tra i due numeri esercita sull'indice di ipernormalità.

RESUMÉ

Cet article groupe les résultats de recherches minutieuses conduites patiemment au cours d'une trentaine d'années, auprès de la Faculté des Sciences statistiques, démographiques et actuarielles de l'Université de Rome et au Comité italien pour l'étude des problèmes de la population, au sujet de la forme des distributions des caractères anthropologiques.

Dans l'intervalle entre les deux guerres mondiales, l'interpolation des distributions des tailles des conscrits italiens, qui s'étaient présentées pour le recrutement et pour lesquelles les relations annuelles du Ministère de la Guerre fournissaient une longue série (de la classe de 1854 à celle de 1920) de données détaillées et comparables, parut un sujet convenable pour les exercices, qui étaient obligatoires pour les étudiants inscrits au cours de Statistique élémentaire de la Faculté de Sciences statistiques, démographiques et actuarielles à l'Université de Rome.

Il s'agissait d'interpoler, sur la distribution des tailles des recrues de chaque circonscription (compartiment ou province ou arrondissement), une courbe normale, en mesurant ensuite le degré d'approximation des données théoriques par rapport aux données réelles et la curtosis ou hypernormalité de la distribution.

On s'est vite rendu compte que, en contraste avec l'opinion courante qui considérerait la taille comme un exemple typique d'un caractère suivant la courbe normale dans sa distribution, celle-ci, par contre, résultait presque toujours hypernormale, les fréquences effectives étant plus élevées de celles théoriques au centre, autour de la moyenne, et aux deux extrémités de la distribution.

La mesure de l'hypernormalité a été fondée sur le rapport entre l'écart quadratique moyen et l'écart simple moyen de la moyenne arithmétique, puisque ces deux constantes peuvent être calculées aisément et très facile aussi est le contrôle des opérations nécessaires pour les obtenir, contrôle qui est indispensable pour les exercices qu'on assigne aux étudiants. L'indice de l'hypernormalité a été déduit de la quantité $I = 2 \left(\frac{{}^2S}{{}^1S} \right)^2$ qui résulte ≥ 3.1416 selon que la distribution est hypernormale, normale ou hyponormale.

Afin de déterminer si l'hypernormalité de la distribution des tailles constituait une qualité spéciale des statistiques italiennes ou au contraire avait une valeur générale, le calcul des indices de l'hypernormalité fût étendu aux données des statistiques, analogues à celles de l'Italie, concernant les recrues françaises et ensuite aux données ayant paru pendant ces années, relatives à la Suisse et à l'Autriche (district de Mistelbach). Les valeurs des indices resultaient d'environ 3.20-3.30.

C'est ainsi que l'hypernormalité de la distribution de la taille des recrues fut confirmée, prenant la valeur d'une régularité statistique. Celle-ci a été énoncée dans une communication présentée à la Société italienne de Statistique dans la séance du mois de juin 1949 et publiée en anglais, avec quelques réductions, l'année suivante dans « *Acta genetica et statistica medica* » de Upsal. Entretemps, les élaborations de nouvelles données concernant les recrues se poursuivaient.

Pour l'Italie, les distributions des tailles relatives à tout le pays ou aux divers compartiments atteignaient le nombre de 166, toutes les distributions en question, à l'exception d'une, résultant hypernormales.

Pour la Suisse, on calculait les indices pour chaque Canton, et 21 sur 25 étaient hypernormaux.

Pour la Suisse et pour l'Italie, on a pu constater que les valeurs des indices relatives à certaines circonscriptions ne différaient pas systématiquement de la valeur de l'indice concernant l'ensemble du territoire et, pour la France et l'Italie, on a pu constater que, analoguement, les indices obtenus pour les différentes années ne différaient pas de l'indice concernant la période toute entière.

Pour la Belgique, on calcula les indices relatifs aux recrues de l'année 1931, pour les distributions non seulement des tailles, mais aussi des périmètres thoraciques et des poids, en obtenant, pour ces derniers caractères, des valeurs (3.41 et respectivement 3.36) sensiblement supérieures à celle des tailles (3.17).

Pour la Belgique, on a pu aussi considérer à part les recrues déclarées aptes au service, en constatant qu'elles montraient des indices d'hypernormalité, pour les trois caractères, inférieurs à ceux de l'ensemble des recrues examinées et cela à cause de la sélection, qui élimine les valeurs extrêmes, les plus basses surtout, mais aussi, dans une certaine mesure, les plus hautes.

En ce qui concerne les tailles, on a pu souligner l'effet de la sélection pour plusieurs autres pays (Italie, Ecosse, Suède, Etats-Unis d'Amérique) en calculant les indices d'hypernormalité pour les distributions des tailles des soldats en service qui résultent fort souvent hyponormales. Une exception est représentée par des distributions des soldats finnois de la fin du XVIII^e siècle ou des premières années du XIX^e, qui présentent des indices d'hypernormalité élevés, mais elle peut être attribuée, tout au moins en partie, au fait que, dans ces temps, la visite médicale n'était pas rigoureuse.

Toutes ces conclusions se fondaient sur les distributions relatives à des sujets du même âge (recrues, conscrits, soldats).

Différents sont les résultats des distributions comprenant toute la population adulte. On a pu recueillir des données concernant le total de la population adulte, mâles et femelles à part, relativement à des villages de la Suisse: Engtlingen et Frutigtal (Oberland), Untertoggenburg (Saint Gall), Schangau (Oberemmenthal) et de l'Allemagne (Genkingen, Brengener Wald et pays voisins), ainsi que pour les Lapons, les Esquimaux, les habitants des îles Faerøer et de certaines populations de la Mélanésie (île de Lir, quatre villages de la côte septentrionale de la Nouvelle-Guinée et quarante-deux localités de la Nouvelle Irlande) et de la Micronésie (Kapingamarangi).

Très souvent, les indices concernant les populations globales résultent hyponormaux, mais en général leurs moyennes résultent hypernormales parce que les valeurs hypernormales présentent des écarts de la valeur moyenne plus forts de celles des indices hyponormaux.

Les recherches relatives aux tailles et, pour certains pays, aux poids et aux périmètres thoraciques ont pu être étendues à plusieurs autres caractères pour les recrues de la Suisse et pour les adultes des susdits villages de la Suisse et de l'Allemagne, ainsi que des îles Faerøer et de la Mélanésie et Micronésie. A l'exception des îles Faerøer, les relèvements de toutes ces données ont été exécutés par Schlaginhaufen ou ses assistants, de sorte qu'elles sont rigoureusement comparables. Ils ont trait à des nombreux caractères quantitatifs (34 pour les recrues suisses et toujours plusieurs dizaines pour les adultes des divers villages), de façon que l'on peut aisément reconnaître une valeur générale, en ce qui concerne les caractères anthropométriques, aux conclusions qu'on en tire.

Pour les recrues, les indices résultent pour la plupart hypernormaux, tandis qu'ils sont très souvent hyponormaux pour les populations adultes.

On a pu comparer les indices obtenus pour 17 caractères de recrues suisses, caractère par caractère, à ceux des adultes appartenant aux deux sexes des villages de la Suisse et des autres pays. Dans la plupart des cas, la différence est évidente; en effet dans 28 cas sur 34 l'hypernormalité résulte supérieure pour les distributions relatives aux recrues. Ce résultat paraît confirmé par les calculs exécutés sur des données fournies par d'autres auteurs au sujet des adultes des îles Faerøer, des Esquimaux, des Samaritains et, en ce qui concerne les caractères du crâne, des anciennes populations des îles Canaries.

La différence entre les distributions des gens du même âge et celles de la population adulte en général, peut s'expliquer par le fait que cette dernière comprend des individus de différents âges, tandis que les valeurs moyennes de plusieurs caractères varient avec l'âge, même parmi les adultes.

Pour certaines populations (Italiens, Lapons, étudiants des collèges américains et certains peuples des îles Canaries), on a pu calculer les indices d'hypernormalité pour divers groupements d'âges.

Dans les populations (des villages, suisses, des villages allemands, de la Mélanésie et de la Micronésie, Turcs, Samaritains, Lapons, étudiants

américains habitants des îles Faeröer et de certaines îles Canaries) pour lesquelles on avait des données séparées pour les deux sexes, on n'a pas pu relever aucune différence systématique entre les indices relatifs aux mâles et ceux relatifs aux femelles.

On a cherché aussi d'établir s'il y a une relation quelconque entre le degré d'hypernormalité des distributions et le rapport de dépendance du caractère en mesure majeure ou mineure de l'héritage ou du milieu (naturel ou social), en arrivant à la conclusion que l'hypernormalité paraît généralement supérieure pour les caractères qui relèvent davantage du milieu.

Enfin, on a pu mettre en évidence sur des données comment les recrues anglaises que, pour les individus ayant des défauts, la mesure de la variabilité (sur la base soit de l'écart simple moyen que de l'écart quadratique moyen) ainsi que en général l'hypernormalité dans la distribution des tailles, des poids et des périmètres thoraciques est plus élevée que pour les individus normaux et augmente avec la gravité de ces défauts, ce qui est en harmonie avec la théorie de Quetelet et des constitutionnalistes italiens, pour lesquels l'homme moyen est aussi le type de l'homme sain.

Une étude à part a été conduite pour les distributions de la réfraction de l'œil, qui présentent une hypernormalité vraiment exceptionnelle en comparaison avec celle constatée pour tous les autres caractères étudiés jusqu'à présent.

Dans ce but on a calculé les indices d'hypernormalité pour les distributions relatives à des patients des cliniques oculistiques ou aussi au personnel médical ou militaire, examinés dans les villes de Tubingue (Allemagne), Chicago (U.S.A.), Goeteborg (Suède), Upsal (Suède), Londres (Angleterre), ainsi qu'à des nègres du Gabon (Afrique) et aux Esquimaux de Anmagssalik (Groenland).

Les données ne sont pas rigoureusement comparables, soit du fait que dans la plupart des relevements, certaines catégories d'yeux ont été éliminées, soit parce que la composition des individus examinés par rapport à leur occupation est très différente, soit à cause des méthodes de relèvement qui n'ont pas été toujours les mêmes. Néanmoins, toutes les distributions démontrent une hypernormalité saillante, en général considérablement supérieure à celle des autres caractères anthropométriques.

Le résultat tout d'abord paraissait singulier d'autant plus que toutes les mesures de l'œil dont la réfraction est la résultante (courbement de la cornée, courbement de la surface externe et interne du cristallin et longueur de l'axe de l'œil (ce dernier à son tour dépendant de la profondeur de la chambre antérieure, de l'épaisseur du cristallin et de la profondeur de la chambre postérieure) résulteraient normalement distribuées. Des recherches plus soignées ont pu toutefois montrer, qu'en réalité, du moins en ce qui concerne la longueur de l'axe, ce résultat relevait du fait que les cas extrêmes avaient été négligés. Lorsque ces derniers furent dûment considérés, la distribution de la longueur de l'axe résulta nettement hypernormale, toujours moins, toutefois, de la réfraction oculaire.

Une représentation graphique de la distribution de la réfraction et

de celle de la longueur de l'axe sur une charte de probabilités a mis en évidence qu'elles peuvent être considérées le résultat de la superposition de trois courbes différentes, dont une relative au type emétrope, c'est à dire normal, et normale aussi (c'est à dire gaussienne) dans sa forme, et les autres deux correspondant aux ectypes de la myopie et de la hypermétropie.

La plus grande hypernormalité de la réfraction de l'oeil par rapport à celle de la longueur de l'axe s'explique par le fait que, jusqu'à ce que les variations de la longueur de l'axe restent entre des limites normales, elles sont compensées par les variations des autres facteurs de la réfraction de l'oeil (courbements de la cornée et des surfaces externe et interne du cristallin), tandis que, lorsque la longueur de l'axe atteint des dimensions supérieures ou inférieures à certaines limites, il n'y a aucune compensation.

Il s'en suit que dans la distribution relative à la réfraction de l'oeil, tandis que les deux ectypes dépendant des longueurs ectypiques de l'axe sont conservés, les valeurs normales de la réfraction se serrent autour de la valeur moyenne et, par conséquent, on relève des valeurs exceptionnellement élevées dans l'indice d'hypernormalité de la réfraction de l'oeil.

A ce point-ci on s'est demandé si une telle explication peut s'étendre aussi à l'hypernormalité qui, bien qu'en mesure inférieure, se vérifie pour les autres caractères anthropométriques.

A ce propos, il y a lieu de faire une distinction entre les caractères fonctionnels et les caractères structurels.

Les caractères fonctionnels dérivent de l'action combinée de plusieurs facteurs qui peuvent se compenser dans leurs effets, de sorte qu'il est possible d'accepter, pour eux aussi, l'explication adoptée pour la réfraction oculaire. On peut considérer que cette conclusion a été confirmée par des contrôles exercés jusqu'à présent sur quelques caractères fonctionnels et qui ont prouvé que l'hypernormalité du caractère dépasse celle des divers facteurs qui le composent.

Par contre, les caractères structurels, tout au moins ceux qui sont linéaires, sont constitués par des éléments pouvant être totalisés et l'on sait bien que les segments d'un même organe ne se compensent pas et ne sont pas même indépendants l'un de l'autre dans leurs variations, mais, dans une certaine mesure, ils sont unis par des liens de solidarité, s'exprimant par des coefficients de corrélation positifs.

C'est ainsi que s'explique le fait que, au contraire de ce qui arrive pour les caractères fonctionnels, les caractères structurels linéaires résultants ne présentent pas une hypernormalité supérieure à celles des facteurs qui les composent.

Les caractères fonctionnels et les caractères structurels linéaires se comportent différemment aussi sous des autres points de vue.

L'hypernormalité est sensiblement plus élevée pour les caractères fonctionnels.

La cogradaution entre hypernormalité et variabilité est positive pour les caractères fonctionnels, tandis que, pour les caractères structurels li-

néaires, quelquefois elle est aussi positive, mais, dans d'autres cas, elle est négative : il arrive même qu'elle soit positive si l'on mesure la variabilité se basant sur l'écart simple moyen et au contraire négative si l'on se base sur l'écart quadratique moyen (standard deviation). Quelque fois il est possible d'expliquer le comportement particulier de la cograduation dans les cas de caractères structurels linéaires ; autrefois on n'y réussit pas.

Pour ce qui concerne les caractères structurels non linéaires (circonférences, volumes, etc.) on a constaté, dans les quelques cas que nous avons pu les examiner, que, sous certains points de vue, ils se comportent comme les caractères fonctionnels, mais, sous des autres, comme des caractères structurels linéaires.

Dans les Appendices, Mlle Anita Trezza démontre que, lorsqu'on additionne deux distributions normales contenant le même nombre de cas, on obtient une distribution hyponormale si les deux distributions données ont la même variabilité et différentes moyennes et, au contraire, une distribution hypernormale si les deux distributions données ont la même moyenne et différentes variabilités, et cela soit que l'on adopte comme mesure de l'hypernormalité l'indice de Pearson ou celui de Gini. Pour le deuxième cas, elle considère aussi l'hypothèse que le nombre des cas contenu dans une distribution est différent de celui contenu dans l'autre et étudie l'influence que le rapport entre les deux nombres exerce sur l'indice d'hypernormalité.

SUMMARY

This article groups the results of minutely detailed research on the form of the distribution of anthropological characters patiently carried out for nearly thirty years at the Faculty of Statistical Demographic and Actuarial Sciences of the University of Rome and the Italian Committee for the study of population problems.

During the interval between the two world wars, the interpolation of the distributions of the stature of conscripts in the army, of which the annual reports of the War Ministry furnished long series (from the class of 1854 to that of 1920) of detailed intercomparable data, seemed a suitable subject for the exercises which were compulsory for the students enrolled in the course of Elementary Statistics of said Faculty.

It was a question of interpolating a normal curve on the distributions of stature of the conscripts in every area (department, province or district), then measuring how close the theoretical data come to the effective data and what is the kurtosis or hypernormality of the distributions.

We quickly realized that, contrary to the current opinion which considered height as the typical example of a character which followed the normal curve in its distribution, this turned out to be almost invariably hypernormal, the effective frequencies being higher than the theoretical

ones for the central values around the average, and for the extreme values of the distribution.

The measurement of hypernormality was based on the ratio between the average square deviation and the average simple deviation from the arithmetical mean in as much as these two constants are easily calculated and it is simple also to check the operations to obtain them, a check which is indispensable for exercises entrusted to students. As index of hypernormality the value of $I = 2 \left(\frac{2S}{1S} \right)^2$ has been taken which is $\gtrless 3.1416$ according to whether the distribution is hypernormal, normal or hyponormal.

In order to decide whether the hypernormality of the stature distribution was a peculiarity only of Italian statistics or whether it had a more general value, the calculation of hypernormality indices was extended to the statistical data of the French conscripts, analogous to those of Italy, and to data which in those years had been published concerning Switzerland and Austria (Mistelbach district). The values of the indices were about 3.2-3.3.

From them the hypernormality was confirmed of the distribution of the height of the conscripts, which thus took on the value of a statistical regularity. This was pointed out in a paper presented to the Italian Statistical Society at the June 1949 meeting and published, somewhat abridged, in English the following year in the «Acta genetica et statistica medica» of Upsal.

In the meantime the processing of new data concerning conscripts was continued.

For Italy, the number of distributions for height, covering the whole country or the individual regions, was increased to 166. All except one were hypernormal.

In Switzerland, the indices were calculated for the individual cantons and, out of 25, 21 were hypernormal.

It was possible to observe that, for Switzerland and Italy, the index values of the different territorial areas did not differ systematically from that of the whole territory. Similarly, it was seen that, for France and Italy, the indices obtained for the single years did not differ from the index for the whole period.

For Belgium, the indices were calculated for the class of 1931, not only on the distribution of height but also for those of chest circumference and weight, obtaining, for these two latter, values (respectively 3.41 and 3.36) appreciably higher than that for height (3.17).

As regards Belgium, the conscripts classified as fit for army service could also be considered separately. They show, for all three characteristics, hypernormal indices much lower than those for the whole number of conscripts examined. This is the effect of the medical selection which eliminates the extreme values, especially the lowest, but also, to a certain degree, the highest.

As regards height, the effect of selection was brought to light in a number of other countries: Italy, Scotland, Finland, Sweden, the United

States of America, calculating the indices for hypernormality for the distribution of stature of the militaries in service, which often turned out to be hyponormal.

An exception was noted in the distribution of stature of Finnish soldiers at the end of the XVIII century and the beginning of the XIXth, where indices of high hypernormality were found, but this may be attributed to the fact that the medical examinations were not very severe as the sources themselves admit.

All these conclusions were based on the distributions covering age contemporaries (conscripts, recruits, soldiers).

The results based on the distribution covering the whole adult population are different.

It was possible to obtain data for the whole adult population, separately for the males and females, for various villages in Switzerland: Engtingen and Frutigtal (Oberland), Untertoggenburg (St.-Gallen), Shangu (Ober-Emmenthal) and in Germany (Genkingen, Bregenzer Wald and nearby places) and also for the Lapps and Eskimos, as well as for the inhabitants of Faeroer Island, and of some islands in Melanesia (Lir, four villages on the north coast of New Guinea and forty-two places in New Ireland) and in Micronesia (Kapingamarangi).

Very often the indices for the total populations are hyponormal, but their averages in general are hypernormal on account of the fact that the individual hypernormal values are much further from the average values than those of the individual hyponormal indices.

It was possible to extend the research concerning height, weight and chest to many other characters on the Swiss conscripts and the adults of the villages mentioned above in Switzerland and Germany as well as of the Faeroer Islands, in Melanesia and Micronesia. Apart from the Faeroer Islands, the taking of all these data was carried out by Schlaginhaufen or his assistants, so that they are strictly comparable. They cover multiple quantitative characteristics (34 for the Swiss conscripts and always several dozen for the adults of the different villages) so that we are justified in attributing, to the conclusions drawn from them, a general value for the anthropometric characters.

The indices covering the conscripts are in a great majority hypernormal, whereas those of the adult population are very often hyponormal.

The indices obtained for 17 characters among the Swiss conscripts could be compared, character by character, with those of the adults of both sexes in the villages in this and the other countries. In the great majority of cases a difference was evident as, 28 times out of 34, the hypernormality was higher for the distributions covering the conscripts. This result seems to be confirmed by the calculations made on the data furnished by other authors for the adults of the Faeroer Islands, the Eskimos and the Samaritans, as well as for the ancient populations of the Canary Isles as regards the cranial characteristics.

The differences between contemporary age groups and whole adult populations is explained by the fact that the latter contains groups of va-

rious ages whereas the average value of many characteristics varies according to the age, including in the adult period of life.

It was also possible to calculate the indices of hypernormality for different age groups in some populations (Italians, Lapps, students of American colleges and the inhabitants of some of the Canary Isles).

No systematic difference has been found between the indices relative to the males and females which could be calculated for various populations (Swiss villages, German villages and those in Melanesia and Micronesia, Turks, Samaritans, Lapps, the inhabitants of the Faeroer Islands, American students and some of the Canary Isles).

It was also sought to establish if there was relation between the degree of hypernormality of the distributions and the dependence of the character in greater or less degree on heredity or environment (natural and social), coming to the conclusion that hypernormality in general seems greater for characteristics that depend more on the environment.

Finally, it was possible to bring to light on the ground of data concerning the English recruits that, for defective individuals, the variability (measured by the mean deviation and by the standard deviation) and, as a rule, also the hypernormality of the distribution of the statures, the weights and the chests was higher than for the normal ones and the higher the more severe the defect, a finding in agreement with Quetelet's theory and that of the Italian constitutionalists according to which the average man is also the type of healthy man.

The distribution of the refraction of the eye was treated apart; this shows an absolutely exceptional hypernormality in comparison with all the other characteristics studied up till now.

To this end, the hypernormality indices for distribution were calculated for patients in eye clinics and also on the medical and military personnel examined at Tübingen (Germany), Chicago (U.S.A.), Goteborg (Sweden), Upsal (Sweden), London (England) and also for negroes in Gabon (Africa) and the Eskimos of Anmagssalik (Greenland).

The data are not strictly comparable, both because in the greater part of the surveys certain categories of eyes were excluded and, on the other hand, the composition of the individuals examined was different according to occupation, and also because the methods of examination were not always the same. In any case, all the distributions show a marked hypernormality, generally much higher than that of the other anthropometric characters.

The results appears singular and all the more so as all the characteristics of the eye upon which refraction depends (curve of the cornea, curves of the external and internal surfaces of the crystalline lens and the length of the axis of the eye, this latter depending, in its turn, on the depth of the anterior chamber, the thickness of the crystalline lens and the depth of the posterior, chamber) were thought to have a normal distribution. More detailed research, however, showed that actually, at least as regards the length of the axis, the result depended on the fact that extreme cases had not been included. When these were properly

considered, the distributions of the length of the axis was found to be clearly hypernormal, always less, however, than the refraction of the eye.

Graphs of the distributions of the refraction and the length of the axis on a chart of probabilities have brought to light how they may plausibly be considered as the result of the superimposition of three component curves: one plotting the emmetrope or normal type — also normal (i.e. Gaussian) in form — the other two corresponding to the two ectypes of myopia and hypermetropia.

The higher hypernormality of the refraction of the eye in comparison with that of the length of the axis may be explained by the fact that, up to the point that the variations of the length of the axis were maintained within normal limits a compensation occurs between them and the variations of the other factors of the refraction of the eye (curves of the cornea and of the external and internal surfaces of the crystalline lens), whereas, when the length of the axis reaches dimensions above or below a given limit, this compensation does not occur. It results that, in the distribution relative to the refraction of the eye, whereas the two ectypes dependent on the ectypical length of the axis are conserved, a marked concentration around the average value occurs in the normal refraction values. From this come the exceptionally high values in the hypernormality index of the refraction of the eye.

At this point it was wondered whether this explanation could also be extended to the hypernormality which, although in lesser degree, is found for the other anthropometric characters.

In this connection, it seems necessary to make a distinction between functional and structural characters.

Functional characters derive from the combined action of various factors which can compensate in their effects so that, for these, the explanation adopted for the refraction of the eye seems possible as well. This conclusion may be regarded as confirmed by some findings on functional characters which have shown the hypernormality of the character to be greater than that of its component factors.

Structural characters, on the other hand, at least those linear in nature, are composed of components which sum up and it is known that the segments of one organ do not compensate nor are they mutually independent in their variations, but, in a certain degree, are linked with bonds of solidarity which are expressed in positive coefficients of correlation.

Thus it is explained that, on the contrary to what occurs with functional characters, structural linear characters do not present a hypernormality greater than that of the component factors.

The functional characteristics and the linear structural characteristics behave differently under other aspects as well.

Hypernormality is notably higher for the functional characteristics.

Cograduation between hypernormality and variability is positive for functional characteristics but, while it may also sometimes be positive for the linear structural characteristics, at other times it is negative and in some cases is negative if the variability is measured by the mean devia-

tion and on the contrary positive if measured by the standard deviation. Sometimes it is possible to explain the particular comportment of the cograduation in the case of linear structural characteristics, at other times it is not.

As regards the non-linear structural characteristics (circumference, volume, etc.) in the cases that it has been possible to examine, them under certain aspects these behave like functional characteristics and under others like linear structural characteristics.

In the Appendices, Miss Anita Trezza shows that, when two normal distributions containing the same number of cases are added together, a hyponormal distribution is obtained if the two given distributions have the same variability and different averages, but, on the contrary, a hypernormal one if the two given distributions have the same average and different variabilities (and this is found whether the Pearson hypernormality index is adopted or Gini's). In the latter case, she also considers the hypothesis of the number of cases contained in a distribution being different from that contained in the other and takes up the question of the influence that the proportion between the two numbers has on the hypernormality index.

Sulla definizione dei concetti di media e di interpolazione ⁽¹⁾

AMATO HERZEL

SOMMARIO: 1. La definizione del Chisini: significato e manchevolezze. 2. Enunciazione di una definizione più generale. 3. Relazioni fra le due definizioni dal punto di vista matematico. 4. Esempificazioni. 5. Considerazioni sui campi di definizione delle medie. 6. Estensione della definizione a distribuzioni di qualunque tipo. 7. Raffronto con la definizione del Chisini. 8. Sulla definizione del concetto di interpolazione.

1. La maggior parte degli statistici italiani accetta oggi la nota definizione del concetto di media, data da O. Chisini nel 1929 (2). Afferma il Chisini: «la ricerca di una media ha lo scopo di semplificare una qualche nostra questione sostituendo, in essa, a due, o più, quantità date una quantità sola che valga a sintetizzarle, senza alterare la visione d'insieme del fenomeno considerato. Pertanto... non ha senso parlare di *media di due* (o più) quantità, ma ha senso parlare di media di esse *all'effetto della valutazione sintetica di un'altra grandezza che ne dipende*». In forma matematica, questo concetto si traduce nella notissima definizione:

Data una funzione

$$y = f(x_1, x_2, \dots, x_n)$$

(1) Lavoro presentato al Seminario di Statistica metodologica dell'Istituto di Statistica della Facoltà di Scienze Statistiche, Demografiche ed Attuariali dell'Università di Roma e discusso nella riunione del 25 gennaio 1958.

(2) O. CHISINI, *Sul concetto di media*, in «Periodico di Matematiche», Serie IV, vol. IX, n. 2.

di un certo numero, n , di variabili indipendenti, $x_1 x_2 \dots x_n$, rappresentanti grandezze omogenee, dicesi media delle $x_1, x_2, \dots x_n$, rispetto alla funzione f quel numero M che sostituito alle $x_1 x_2 \dots x_n$ dà il medesimo valore per la f che le $x_1 x_2 \dots x_n$ stesse, cioè quel numero M tale che :

$$f(M, M, \dots, M) = f(x_1 x_2, \dots, x_n) \quad [\text{I}]$$

Dalle parole citate, come anche da tutto il contesto dell'articolo, risulta evidente che la definizione data si propone di spiegare la vera natura delle medie, e non già di fornirne una giustificazione puramente formale. Di conseguenza la funzione f che ne costituisce il fulcro, deve esprimere un'effettiva relazione fra le quantità date e *un'altra* grandezza (nei noti esempi addotti dal Chisini, le quantità date sono delle velocità e l'altra grandezza è costituita dal tempo impiegato o dal carburante consumato); non sarebbe quindi corretto servirsi della definizione per spiegare, per esempio, l'impiego della mediana, ponendo la f uguale al termine centrale delle quantità, $x_1, x_2 \dots x_n$ disposte in ordine crescente. Un simile procedimento sarebbe tautologico e privo di valore conoscitivo; adottandolo si farebbe in sostanza quell'« opera puramente formale ed antifilosofica » che il Chisini giustamente rimprovera a coloro che si limitano a « definire le singole specie di medie che si incontrano abitualmente ».

Ciò premesso, è però facile rendersi conto che la definizione di CHISINI non si applica ad alcune importanti specie di medie, largamente impiegate dagli statistici o comunque di interesse teorico. Già abbiamo accennato alla mediana, ma lo stesso dicasi per esempio della moda e dei cosiddetti « centri » di ordine r (cioè le quantità che rendono minima la somma delle potenze r esime dei valori assoluti degli scarti) o anche di medie di distribuzioni particolari, e precisamente le medie delle serie cicliche. Come si vede, la definizione del Chisini risulta inapplicabile non soltanto alle cosiddette medie di posizione, ma anche ad alcune medie analitiche.

Questa constatazione, in verità non nuova, ma finora trascurata, a torto, ci sembra, deve indurre ad un riesame critico di tutta la definizione, diretto a scoprire ed eliminare la causa di queste lacune.

2. Si possono prendere le mosse da un esempio semplicissimo, simile a quelli usati dal Chisini. Un treno parte dalla località A, passa per la località B e termina la sua corsa in C. Sia a la distanza (ferroviaria) fra A e B e b quella fra B e C. Il treno percorre il tratto da A a B alla velocità di v_a chilometri orari, e quello da B a C alla velocità di v_b chilometri. L'amministrazione ferroviaria desidera ora sostituire questo treno con un altro, che viaggi a velocità uniforme.

Se questa sostituzione deve lasciare inalterata la durata del percorso complessivo, la velocità media ricercata sarà data ovviamente dalla solita formula :

$$v_m = \frac{a + b}{\frac{a}{v_a} + \frac{b}{v_b}}$$

cioè dalla media armonica ponderata. Ciò rientra perfettamente nello schema del Chisini.

Ma, evidentemente la sostituzione di una velocità uniforme alle due velocità date può essere operata anche con criteri del tutto diversi, adattati alle esigenze più svariate. Per esempio, può essere utile fare sì che l'orario venga alterato il meno possibile dalla sostituzione, e precisamente può riuscire conveniente non modificare l'orario del passaggio del treno per la località B e determinare v_m in modo che l'ora della partenza da A e quella dell'arrivo a C subiscano complessivamente il minore spostamento possibile. In tal caso, v_m sarà data da quel valore di v che rende minima la funzione :

$$\left| \frac{a}{v_a} - \frac{a}{v} \right| + \left| \frac{b}{v_b} - \frac{b}{v} \right|$$

Si avrà quindi $v_m = v_a$ se $a > b$ e $v_m = v_b$ se $b > a$. In questo caso siamo fuori dello schema del CHISINI, in quanto la media cercata viene determinata non da un'equazione, ma da una condizione di minimo.

Da questo, e dai numerosissimi altri esempi che si potrebbero ricavare da tutti i campi della vita quotidiana, è possibile trarre alcune conclusioni che ci permettono di colmare le lacune insite nella definizione del Chisini :

1) Non è esatto, in linea generale, che le medie servano soltanto alla stima di altre grandezze. L'esigenza di sintesi, che si estrinseca nella ricerca di una media, può essere determinata da altre circostanze, di natura svariaticissima. Il Chisini, in sostanza ha prospettato solo un caso particolare, sia pure di grande importanza.

2) Se la circostanza che ci interessa quando ci accingiamo a determinare una media non consiste nella stima di una quantità, non potremo esigere che la media non alteri « la visione d'insieme del fenomeno considerato ». In generale, una certa alterazione del fenomeno sarà inevitabile ; fra tutti i valori che a priori potremo attribuire alla media, sceglieremo quindi quello che rende minima tale alterazione, o scostamento — che in sostanza costituisce il costo della « operazione media » — ai fini della circostanza che ci interessa e dello scopo che ci proponiamo di raggiungere.

Perchè il problema sia determinato, occorre supporre l'esistenza di una funzione g che misuri appunto l'alterazione che verrebbe apportata al fenomeno considerato dai vari valori che la media potrebbe assumere. La forma di tale funzione dipenderà ovviamente dalla natura della questione che vogliamo semplificare o risolvere con l'uso della media e dalle particolari esigenze di cui vogliamo o dobbiamo tenere conto. Questa funzione determina la media, che è data appunto da quel valore (di un insieme fisso) che la rende minima. Le forme assunte dalla funzione possono essere evidentemente diversissime ; essa tuttavia deve sempre soddisfare ad una condizione ovvia : se tutte le quantità date sono uguali, ovvero (in quanto appare più conveniente, trattandosi di un problema prevalentemente statistico, ragionare in termini di modalità e frequenze) quando il fenomeno presenta una sola modalità, esso non si altera, quando viene sintetizzato dalla stessa modalità, e quindi in tale caso la funzione g deve assumere il suo minimo valore quando la media viene a coincidere con la modalità data.

Infine riteniamo anche noi, col Gini (1) che sia preferibile riservare la qualifica di medie ai valori minimizzanti la g , che sono compresi nell'intervallo che ha per estremi la minima e la massima delle modalità x_1, x_2, \dots, x_k e di designare col termine di «adeguati numerici» i valori che eventualmente minimizzassero la g senza soddisfare a tale condizione.

In base a quanto esposto, si giunge alla seguente *definizione*:
« Sia data una funzione:

$$y = g(m; x_1, x_2, \dots, x_k, n_1, n_2, \dots, n_k)$$

dove le x_1, x_2, \dots, x_k rappresentano modalità di un carattere quantitativo o comunque espresso quantitativamente, e le n_1, n_2, \dots, n_k le frequenze assolute che ad esse competono, mentre la m è una variabile reale definita in un dato insieme e sia verificata per ogni costante a compresa nell'insieme di definizione delle x_1, x_2, \dots, x_k la relazione:

$$\min. g(m; a, n) = g(a; a, n) \quad [2]$$

(dove si è posto $\sum_{i=1}^k n_i = n$). Se al variare della variabile m nel suo campo di definizione la funzione

$$y = g(m; x_1, x_2, \dots, x_k, n_1, n_2, \dots, n_k)$$

assume il suo minimo valore quando la m assume il valore M , e se inoltre detto valore non è minore della più piccola nè maggiore della più grande delle modalità x_1, x_2, \dots, x_k , la quantità M dicesi media della distribuzione $(x_1, x_2, \dots, x_k, n_1, n_2, \dots, n_k)$ rispetto alla funzione g . Se quest'ultima condizione non è soddisfatta, la quantità M dicesi adeguato numerico della stessa distribuzione (2).»

3. Già dalle considerazioni che l'hanno preceduta, risulta che la nostra definizione comprende quella di Chisini come caso par-

(1) C. GINI, *L'evoluzione del concetto di media*, in «Metron», vol. XVI, n. 3-4 pag. 19.

(2) Può essere interessante osservare che alcuni dei concetti esposti sono già impliciti nella nozione di «media condizionata» introdotta nella metodologia statistica da S. VIANELLI (V. *Lineamenti di Metodologia Statistica per l'Analisi Economica delle Serie Storiche*, Bologna, 1946). Difatti, le «medie condizionate» possono essere definite non solo da equazioni, ma anche da condizioni di minimo.

ticolare. Ciò viene confermato per via matematica nel seguente modo semplicissimo.

La relazione [I], si può scrivere come segue :

$$f(M; n) = f(x_1, x_2, \dots, x_k, n_1, n_2, \dots, n_k) \quad [I, a]$$

Se si pone ora :

$$\begin{aligned} g(m; x_1, x_2, \dots, x_k, n_1, n_2, \dots, n_k) = \\ = |f(m; n) - f(x_1, x_2, \dots, x_k, n_1, n_2, \dots, n_k)| \end{aligned}$$

la [2] è evidentemente soddisfatta, e la g assume il suo minimo valore per $m = M$.

Viceversa, nel caso che la g permetta di ricavare la M come funzione esplicita delle $x_1, x_2, \dots, x_k, n_1, n_2, \dots, n_k$, che si possa cioè scrivere :

$$M = h(x_1, x_2, \dots, x_k, n_1, n_2, \dots, n_k),$$

dalla (2) si trae :

$$M = h(x_1, x_2, \dots, x_k, n_1, n_2, \dots, n_k) = h(M, n),$$

cioè una relazione del tipo della [I, a], la quale, però, come si è detto prima, può avere in molti casi carattere puramente formale.

4. A titolo di esempio scriviamo ora alcune delle espressioni che la g può assumere in relazione alle medie più comuni :

Media aritmetica semplice :

$$g = \left| \sum_{i=1}^k x_i n_i - n m \right|,$$

oppure :

$$g = \sum_{i=1}^k (x_i - m)^2 n_i$$

Media geometrica semplice :

$$g = \left| \sum_{i=1}^k x_i^{n_i} - m^n \right|$$

oppure :

$$g = \sum_{i=1}^k (\log x_i - \log m)^2 n_i$$

Mediana :

$$g = \sum_{i=1}^k |x_i - m| n_i$$

« Centro » di ordine r :

$$g = \sum_{i=1}^k |x_i - m|^r n_i$$

Moda :

$$g = \sum_{i=1}^k p_i n_i$$

dove :

$$p_i = 1 \text{ per } x_i \neq m$$

$$p_i = 0 \text{ per } x_i = m$$

Valore pozicre :

$$g = \sum_{i=1}^k q_i n_i,$$

dove :

$$q_i = x_i \quad \text{per } x_i \neq m$$

$$q_i = 0 \quad \text{per } x_i = m$$

Media aritmetica della serie ciclica :

$$g = \sum_{i=1}^k s_i^2 n_i,$$

dove :

$$s_i = |x_i - m| \quad \text{per} \quad |x_i - m| < \frac{k}{2}$$

$$s_i = |x_i - m| - \frac{k}{2} \quad \text{per} \quad |x_i - m| > \frac{k}{2}$$

$$\text{e } x_i = i \quad (i = 1, 2, \dots, k).$$

Mediana della serie ciclica :

$$g = \sum_{i=1}^k s_i n_i$$

e così via. Questi esempi non apportano nulla di nuovo, in quanto non fanno che esprimere proprietà ben note delle singole medie, ed in alcuni casi (« Centri » e medie delle serie cicliche) riproducono semplicemente le definizioni abituali. Ciò sta ad indicare che la funzione g , nelle sue varie espressioni, corrisponde effettivamente ad una caratteristica intrinseca del concetto di media. Si tenga presente, per altro, che ad ogni media possono corrispondere infinite altre funzioni, di forma svariaticissima, che abbiano in comune la sola proprietà di essere rese minime dallo stesso valore della variabile m . Ciò spiega perchè il numero dei tipi di medie abitualmente usati sia relativamente ridotto.

5. È interessante rilevare un'altra differenza tra la nuova definizione e quella data dal Chisini. Se, come generalmente accade la [1] ammette un'unica soluzione, una volta assegnate le quantità x_i , la M risulta determinata, e non è quindi possibile imporre restrizioni circa l'insieme dei valori che essa può assumere. Diversamente stanno invece le cose quando si pone a base della definizione la funzione g . La M infatti, non deve in tal caso soddisfare un'equazione, ma una condizione di minimo, ed è quindi

possibile delimitare, comunque fissate le quantità x_i , l'insieme dei valori che la m quindi la M , può assumere, purchè rimanga valida la condizione [2]. Così per esempio a seconda dello scopo cui la media deve servire si può imporre alla media stessa di coincidere con una delle quantità x_i , il che escluderebbe le cosiddette medie di conto; di essere un numero intero positivo ecc. È evidente che questa particolarità risponde ad una esigenza che si fa sentire in molti problemi pratici, in quanto ad esempio i numeri frazionari, irrazionali o negativi per molte questioni sono privi di significato. In pratica si usa in questi casi ricorrere a valori approssimati di quelli che si ottengono risolvendo la (1), ma nella definizione di Chisini non vi è nulla che possa giustificare tale procedimento. Nell'ambito della nuova definizione proposta invece, la soluzione diventa assolutamente rigorosa.

6. È facile adattare la definizione sopra data, valida per le distribuzioni discontinue, al caso, più generale, che si abbia a che fare con distribuzioni di qualunque tipo. Basta sostituire alla funzione g il funzionale $G(y; x, F(x))$, dove $F(x)$ è la funzione di ripartizione della variabile x , e sostituire alla seconda la condizione seguente:

$$\min. G(y; a, \varepsilon(x - a)) = G(a; a, \varepsilon(x - a)) \quad [2, a]$$

dove $\varepsilon(z)$ è la funzione di ripartizione di una distribuzione concentrata nel punto di ascissa $z = 0$ (1). Piuttosto ovvi sono anche gli adattamenti che bisogna apportare agli esempi sopra citati e pertanto non ci soffermiamo su questo punto.

7. Si è già visto che la definizione di Chisini non si applica ad un gruppo di medie che rientrano invece nella definizione proposta. Altre medie, invece, rientrano in entrambe le definizioni e conviene scegliere una di queste ultime per illustrare la differenza fra le due formulazioni. Si prenda a questo scopo la media

(1) Questa funzione, per alcuni autori che postulano le funzioni di ripartizione continue a sinistra, è uguale a 0 per $z \leq 0$ e uguale a 1 per $z > 0$; per altri che le considerano continue a destra, come ci sembra conveniente, è uguale a 0 per $z < 0$ e uguale a 1 per $z \geq 0$. Altri ancora pongono: $\varepsilon(z) = 0$ per $z < 0$; $\varepsilon(z) = \frac{1}{2}$ per $z = 0$, $\varepsilon(z) = 1$ per $z > 0$.

aritmetica. Chiediamoci quindi : quando si usa la media aritmetica ? A questa domanda, la definizione di CHISINI risponde : quando si tratta di lasciare inalterata la somma dei termini. La definizione proposta, invece, fornisce la seguente risposta : quando questa media rende minima l'alterazione del fenomeno rispetto allo scopo che ci si prefigge. Si vede immediatamente che quando la prima risposta è esatta, è esatta anche la seconda. Ma è, poi, sempre esatta questa risposta ? Per convincersi che non è così, basta un esempio classico : il problema delle misure ripetute. Un oggetto è stato misurato con uguale accuratezza, e nelle stesse condizioni, n volte ; le misure abbiano dato n_1 volte il risultato x_1 , n_2 volte il risultato x_2 , ... n_k volte il risultato x_k . Si tratta di sintetizzare queste quantità per ricavarne se possibile, la « vera » misura dell'oggetto. L'alterazione del fenomeno è misurata qui dalla probabilità di accettare come vera una misura sbagliata. Si conviene generalmente di scegliere, come misura « più plausibile », la quantità che rende minima la somma dei quadrati degli scarti, cioè la media aritmetica delle misure eseguite (1). Questo caso rientra perfettamente nella definizione proposta, mentre contraddice alla definizione di Chisini, in quanto l'eguaglianza $\sum_{i=1}^k x_i = n M$ è del tutto irrilevante agli effetti del problema. Ciò dimostra che anche per quei tipi di medie ai quali si può talvolta applicare soddisfacentemente la definizione di Chisini, la nuova formulazione risulta preferibile.

Per concludere sulle analogie e le differenze fra le definizioni del Chisini e quella da noi proposta, riteniamo di poter affermare che quest'ultima pur avendo carattere di maggiore generalità, conserva inalterato il pregio principale della prima definizione, in quanto, al pari di quella, mette l'accento sulla relatività del concetto di media, cioè sul fatto che ogni media trae la sua giustificazione soltanto dal particolare problema che essa è chiamata a semplificare e risolvere. Mentre però il rapporto fra la media e le circostanze che la determinano si presenta nella definizione del Chisini in una forma certo suggestiva, ma in so-

(1) Sulle giustificazioni di tale criterio, vedasi, p. es., G. CASTELNUOVO, *Calcolo delle probabilità*. vol. I, Bologna. 1948, pagg. 232 e sgg.

stanza alquanto rigida e meccanica, nella definizione da noi proposta, tale rapporto, ci sembra, risulta più duttile e più vario, e quindi più soddisfacente.

Osserviamo ancora che si può porre (almeno formalmente) :

$$\varphi(m; x_1, x_2, \dots, x_k, n_1, n_2, \dots, n_k) = g - \min g$$

e allora la media viene determinata dalla funzione implicita :

$$\varphi(m; x_1, x_2, \dots, x_k; n_1, n_2, \dots, n_k) = 0$$

il che conferma la maggiore generalità della nostra definizione rispetto a quella di Chisini, sintetizzata dalla [1,a], in quanto è sempre possibile esprimere una funzione esplicita sotto forma di funzione implicita, mentre non è vero l'inverso.

Non sarà forse del tutto superflua un'altra osservazione. La nuova formulazione è più generale di quella di Chisini e presenta quindi in maggiore misura il pericolo di dar luogo alla costruzione di medie artificiali, prive di significato reale. Ma è chiaro che una definizione generale, come quella proposta, non può che stabilire criteri formali, mentre ogni media, per poter essere considerata tale, deve avere una sua giustificazione intrinseca, deve cioè corrispondere ad un'esigenza effettiva, di carattere teorico o pratico.

8. Se una media sintetizza una distribuzione data, una curva interpolatrice sintetizza del pari una distribuzione data o una funzione empirica ad essa collegata, con un insieme di parametri di una funzione, in genere continua. Da questa constatazione discende che la media può essere considerata un parametro di una curva interpolatrice e di conseguenza tutta la teoria della media potrebbe essere svolta nel quadro della teoria dell'interpolazione. Riteniamo che una simile impostazione potrebbe risultare utile, in quanto permetterebbe una visione più organica di importanti settori della metodologia statistica che finora sono sempre stati trattati in capitoli staccati e da punti di vista differenti.

La definizione del concetto di media da noi proposta si adatta a nostro parere particolarmente bene ad una simile impostazione. È noto infatti che la determinazione dei parametri delle curve

interpolatrici risponde allo stesso criterio di minimo scostamento che sta alla base della nostra definizione. È facile mettere in evidenza questa corrispondenza anche sotto l'aspetto formale.

Sia $f(x)$ la funzione empirica che si vuole interpolare. $\Phi(x; y_1, y_2, \dots, y_k)$ una famiglia di curve (es. parabole, curve esponenziali ecc.), i cui parametri y_1, y_2, \dots, y_k sono da determinarsi; sia dato il funzionale:

$$G[\Phi(x; y_1, y_2, \dots, y_k); x, f(x)]$$

che soddisfa alla condizione:

$$\begin{aligned} \min. G[\Phi(x; y_1, y_2, \dots, y_k); x, \Phi(x; a_1, a_2, \dots, a_k)] = \\ = G[\Phi(x; a_1, a_2, \dots, a_k); x, \Phi(x, a_1, a_2, \dots, a_k)] \end{aligned} \quad (2, b)$$

allora, fissata comunque la funzione $f(x)$ dette $\bar{y}_1, \bar{y}_2, \dots, \bar{y}_k$ le coordinate del punto nello spazio a k dimensioni nel quale la G assume il suo minimo valore al variare della variabile k pla $(\bar{y}_1, \bar{y}_2, \dots, \bar{y}_k)$ in un insieme determinato, la $\Phi(x, \bar{y}_1, \bar{y}_2, \dots, \bar{y}_k)$ si dirà curva interpolatrice della $f(x)$ rispetto alla G .

La (2,b) esprime una condizione ovvia, che può essere così formulata: se alla funzione da interpolare, $f(x)$ si sostituisce una funzione della famiglia Φ , la funzione interpolatrice deve identificarsi con tale funzione.

A prima vista potrebbe sembrare che vi sia una differenza radicale fra questa definizione, nella quale figura la famiglia di funzioni Φ , e quella del concetto di media, prima enunciata, in cui non vi è nulla di analogo. Questa differenza trova la sua spiegazione, oltre che nella maggiore generalità del concetto di interpolazione, nel fatto che la media deve essere assimilata non alla curva interpolatrice, ma ai parametri di questa.

Ciò premesso, risulta tuttavia evidente il nesso fra le due definizioni: basta, infatti, considerare che — salvo le differenze di simbologia — si rientra nel caso della media se la funzione da interpolare è, per esempio, una funzione di ripartizione (ma si giunge allo stesso risultato anche con una funzione di densità, una curva di concentrazione, di graduazione, ecc.) e la funzione interpolatrice è la $\varepsilon(x-m)$, precedentemente definita.

In conclusione, appare lecito, come già avevamo accennato, considerare la media come parametro di particolari funzioni interpolatrici (ad un solo parametro).

Questa impostazione, che mette in rilievo il legame fra curve interpolatrici e medie, potrebbe probabilmente dare impulso a nuove fruttuose ricerche. Mentre infatti le proprietà delle medie sono state fatte oggetto di numerosi studi a carattere sistematico, che hanno anche condotto, come è noto, a diversi tipi di classificazioni, tutto, o quasi rimane ancora da fare in questo campo per quanto riguarda le curve interpolatrici e i metodi di interpolazione. Sarebbe utile, ci sembra, cominciare a colmare questa lacuna.

RIASSUNTO

Partendo dalla considerazione che la media di una distribuzione data deve servire a sintetizzare la distribuzione stessa e che tale sintesi deve essere effettuato in modo da alterare il meno possibile, tenuto conto dello scopo concreto che ci si propone di raggiungere, la visione d'insieme del fenomeno descritto dalla distribuzione, si giunge a definire la media come quel valore che rende minima una funzione data. Tale funzione misura appunto l'alterazione della visione del fenomeno. In questa ampia definizione rientrano tutti i tipi di medie impiegate dagli statistici, come la mediana, la moda, ecc., e vi rientrano in particolare tutte le medie alle quali si applica la nota formulazione del Chisini. Vengono infine messi in rilievo gli stretti legami fra il concetto di media, così inteso, e il concetto di interpolazione.

GIUSEPPE LETI

Nuovi tipi di distanze fra insiemi di punti e loro applicazioni alla statistica

SOMMARIO.

L'Autore introduce due nuove definizioni di distanze fra due insiemi, costituiti da un numero finito di elementi, in base alle quali dà un'impostazione unitaria al problema della dissomiglianza ed a quello di Hitchcock-Koopmans, mostrando come le soluzioni dei due problemi siano casi particolari delle distanze definite. Dà inoltre un'estensione degli indici di dissomiglianza al caso delle distribuzioni multiple.

I. — INTRODUZIONE.

Nel 1914 il Gini introduceva nella statistica il concetto di dissomiglianza fra due seriazioni e gli indici atti a misurarla. Ognuno di tali indici può interpretarsi ⁽¹⁾ come misura della distanza fra due insiemi costituiti da un numero finito di punti, appartenenti ad una retta, e non è altro che il minimo delle medie degli spostamenti a cui debbono essere sottoposti i punti di uno dei due insiemi affinché questo si sovrapponga all'altro ⁽²⁾.

⁽¹⁾ CASTELLANO, V. - *Sulle relazioni fra variabili statistiche*. Appunti di statistica metodologica. Facoltà di scienze statistiche. Roma, 1954-55. POMPILJ, G. - *La mutabilità dell'universo dei campioni*. Scritti matematici in onore di F. Sibirani. Bologna, Zuffi, 1957. LANDENNA, G. — *Osservazioni sulla commessione*. Statistica, anno XVII, n. 1, 1957.

⁽²⁾ P. LEVY, in una nota intitolata: *Distance de deux variables aléatoires et distance de deux lois de probabilité*, contenuta nel volume: FRÉCHET, M. — *Généralités sur les Probabilités. Éléments aléatoires*. Vol. I, Gauthier-Villars, Paris, 1950, ha definito, usando la terminologia delle variabili aleatorie, una distanza fra due distribuzioni statistiche semplici, distanza che nel caso di distribuzioni non è altro che l'indice semplice di dissomiglianza del Gini. Non perviene però ad una formula di tale distanza, anzi dichiara esplicitamente che essa presenta l'inconveniente di poter essere molto difficile da determinare.

Estendendo tale concetto alla misura della distanza fra due insiemi, costituiti da un numero finito di punti, che appartengono ad uno spazio di dimensione superiore, si perviene ad una definizione di distanza, che è diversa da quella che usualmente è data in matematica, ma che, a differenza di quella, è applicabile anche a problemi concreti: infatti è un suo caso particolare la soluzione del problema di Hitchcock-Koopmans, detto anche dei trasporti, che è un problema di programmazione lineare.

I nuovi tipi di distanza consentono inoltre di estendere al caso di più dimensioni gli indici di dissomiglianza fra distribuzioni semplici.

2. — DISTANZA FRA DUE PUNTI.

Ricordiamo che uno spazio si dice *distanziale* o *metrico* se, per ogni coppia di punti P e Q di esso, è definito un numero reale $d(P, Q)$, detto distanza di P da Q , tale che:

- a) sia $d(P, Q) = 0$, se e solo se P e Q coincidono;
- b) presi comunque tre punti P, Q , ed R risulti

$$d(P, Q) \leq d(P, R) + d(R, Q).$$

Di solito alle proprietà a) e b) si sogliono aggiungere le seguenti (che sono conseguenza delle precedenti):

c) $d(P, Q) \geq 0$, cioè la distanza di due punti non sia mai negativa;

d) $d(P, Q) = d(Q, P)$, cioè la distanza di P da Q sia uguale a quella di Q da P .

Consideriamo, per es., la distanza fra due punti di uno spazio numerico, ogni punto del quale è una r -pla di numeri reali; essa può essere definita in vari modi ⁽³⁾: indicati con p_i e q_i , ($i = 1, 2, \dots, r$), rispettivamente le coordinate di P e Q , la distanza può essere, per es., quella euclidea, data dalla formula

$$d(P, Q) = \sqrt{\sum_1^r (p_i - q_i)^2}, \quad [1]$$

oppure quella definita dall'espressione

⁽³⁾ Si veda, per es., BUSEMAN, H. - KELLY, P. J., *Projective Geometry and Projective Metrics*. Academic Press. Inc. Pub. New York 1953, pagine 105-106.

$$d(P, Q) = \sum_1^r |p_i - q_i| \quad [2]$$

od anche quella espressa da

$$d(P, Q) = \max |p_i - q_i| \quad (i = 1, 2, \dots, r). \quad [3]$$

Ricorrendo al concetto di *spostamento*, può attribuirsi alla distanza [1], ed in senso lato anche alla [2], una rappresentazione cinematica, interpretandola come lo *spostamento minimo* che un punto mobile coincidente con P deve effettuare per portarsi su Q e viceversa.

3. — TIPO DI DISTANZA FRA DUE INSIEMI DI PUNTI DEFINITO NELLA MATEMATICA.

Si abbiano due insiemi E ed I di punti tali che per ogni punto di E sia definita la distanza da ogni punto di I e viceversa. Solitamente si definisce distanza fra E e I l'estremo inferiore dell'insieme numerico costituito dalle distanze $d(P, Q)$, al variare comunque di P in E e di Q in I . Se i due insiemi sono costituiti entrambi da un numero finito di punti, la distanza viene ad essere il minimo delle distanze dei punti, di E da quelli di I .

La definizione ora data non soddisfa, qualora si sostituisca la parola « punti » con la parola « insiemi », alla proprietà *a*) di cui gode invece la distanza fra due punti; infatti se due insiemi hanno un punto in comune, la loro distanza è zero, ma non è detto che i due insiemi coincidano; non solo, ma se almeno uno dei due insiemi è costituito da infiniti punti, la distanza può essere anche zero, senza che i due insiemi abbiano un punto in comune.

Se i due insiemi sono costituiti da un numero finito di punti, la distanza, sopra definita, può interpretarsi cinematicamente come lo spostamento minimo che si deve effettuare perchè i due insiemi abbiano almeno un punto in comune. È evidente come tale interpretazione sia di scarso significato, *soprattutto nella statistica*.

4. — NUOVI TIPI DI DISTANZA FRA DUE INSIEMI DI PUNTI.

Consideriamo due insiemi E ed I costituiti ciascuno da un numero n (finito) di punti tali che, per ogni punto di E , sia definita la distanza da ogni punto di I e viceversa; esaminiamo dapprima il caso che in ciascun insieme i punti siano distinti a due a due.

Indichiamo con $E_1, E_2, \dots, E_h, \dots, E_n$ gli n punti di E ; con $I_1, I_2, \dots, I_k, \dots, I_n$ i punti di I e con d_{hk} la distanza $d(E_h, I_k)$. Nell'insieme di tutte le coppie (E_h, I_k) consideriamo tutti i possibili gruppi di n coppie, nei quali ogni punto di E compaia una ed una sola volta in una coppia ed ogni punto di I una ed una sola volta in una coppia. Il numero dei gruppi è $n!$ in quanto ogni gruppo può ottenersi associando ogni elemento della permutazione E_1, E_2, \dots, E_n , con l'elemento che occupa lo stesso posto in una delle $n!$ permutazioni degli I_1, I_2, \dots, I_n . Ciascuno dei suddetti gruppi individua una corrispondenza biunivoca senza eccezione fra i punti di E e quelli di I .

Indichiamo con $G^{(1)}, G^{(2)}, G^{(3)}, \dots, G^{(n)}$ gli $n!$ gruppi, con $G^{(j)}$ uno qualunque di essi, con $I_h^{(j)}$ l'elemento di I che in $G^{(j)}$ è associato all'elemento E_h di E e con $d_h^{(j)}$ la distanza $d(E_h, I_h^{(j)})$, e consideriamo i valori

$${}^1D^{(j)} = \sum_h d_h^{(j)} \quad [4]$$

e

$${}^2D^{(j)} = \sqrt{\frac{\sum_h (d_h^{(j)})^2}{n}} \quad [5]$$

che in generale variano al variare di j .

Il minimo di ${}^1D^{(j)}$, che non è negativo e che potrebbe essere assunto anche in più di un gruppo, lo diciamo *distanza del primo ordine fra E ed I* e lo indichiamo con ${}^1d(E, I)$; il minimo di ${}^2D^{(j)}$, che similmente non è negativo e che potrebbe essere assunto in più di un gruppo, lo denominiamo *distanza del secondo ordine fra E ed I* e lo indichiamo con ${}^2d(E, I)$.

Consideriamo, per es., i due insiemi E ed I , i cui punti hanno le reciproche distanze indicate nella seguente tabella,

nella quale il numero che occupa l' h -esima riga e la k -esima colonna è la distanza $d(E_h, I_k)$:

$\begin{array}{c} \diagdown \\ E \end{array} \begin{array}{c} I \end{array}$	I_1	I_2	I_3
E_1	4	3	5
E_2	1	2	3
E_3	4	1	4

Vi sono 6 possibili gruppi di coppie:

$$G^{(1)}: (E_1, I_1), (E_2, I_2), (E_3, I_3)$$

$$G^{(2)}: (E_1, I_2), (E_2, I_3), (E_3, I_1)$$

$$G^{(3)}: (E_1, I_3), (E_2, I_1), (E_3, I_2)$$

$$G^{(4)}: (E_1, I_3), (E_2, I_2), (E_3, I_1)$$

$$G^{(5)}: (E_1, I_2), (E_2, I_1), (E_3, I_3)$$

$$G^{(6)}: (E_1, I_1), (E_2, I_3), (E_3, I_2)$$

in corrispondenza dei quali si hanno i valori:

$${}^1D^{(1)} = \frac{10}{3}; \quad {}^2D^{(1)} = \sqrt{\frac{36}{3}}$$

$${}^1D^{(2)} = \frac{10}{3}; \quad {}^2D^{(2)} = \sqrt{\frac{34}{3}}$$

$${}^1D^{(3)} = \frac{7}{3}; \quad {}^2D^{(3)} = \sqrt{\frac{27}{3}}$$

$${}^1D^{(4)} = \frac{11}{3}; \quad {}^2D^{(4)} = \sqrt{\frac{45}{3}}$$

$${}^1D^{(5)} = \frac{8}{3}; \quad {}^2D^{(5)} = \sqrt{\frac{26}{3}}$$

$${}^1D^{(6)} = \frac{8}{3}; \quad {}^2D^{(6)} = \sqrt{\frac{26}{3}}.$$

Si ha perciò

$${}^1d(E, I) = {}^1D^{(3)} = \frac{7}{3}; \quad {}^2d(E, I) = {}^2D^{(5)} = {}^2D^{(6)} = \sqrt{\frac{26}{3}}$$

e quindi mentre la distanza del 1° ordine è assunta in $G^{(3)}$ quella del 2° ordine lo è in $G^{(5)}$ e in $G^{(6)}$.

Si vede perciò che le due distanze possono essere assunte in gruppi fra loro diversi.

5. — PROPRIETÀ DELLE DISTANZE DEL PRIMO E DEL SECONDO ORDINE.

Le distanze, così definite, godono delle proprietà a cui soddisfa la distanza di due punti, quando si sostituisca la parola « punti » con « insiemi ». Infatti :

a') se i due insiemi coincidono, ossia se ogni punto di E coincide con un punto di I e viceversa, considerando il gruppo $G^{(j)}$ delle coppie in ognuna delle quali $E_h \equiv I_h^{(j)}$, si ha $d(E_h, I_h^{(j)}) = 0$ per ogni h . Da ciò segue sia ${}^1D^{(j)} = \sum_h d(E_h, I_h^{(j)}) = 0$ e perciò ${}^1d(E, I) = 0$, sia

$$x {}^2D^{(j)} = \sqrt{\frac{\sum_h [d(E_h, I_h^{(j)})]^2}{n}} = 0$$

e perciò ${}^2d(E, I) = 0$. Viceversa se è ${}^1d(E, I) = 0$ oppure ${}^2d(E, I) = 0$, dovrà trovarsi un gruppo $G^{(j)}$, nel quale si abbia per ogni coppia $(E_h, I_h^{(j)})$ di tale associazione, $d(E_h, I_h^{(j)}) = 0$, ossia, per la proprietà *a)* del n. 2, $E_h \equiv I_h^{(j)}$, per ogni h ;

b') presi comunque tre insiemi E, I, J , di n punti ciascuno, per i quali siano definite le distanze fra i punti di ogni insieme da quelli di ogni altro, si ha :

$${}^1d(E, I) \leq {}^1d(E, J) + {}^1d(J, I) \quad [6]$$

$${}^2d(E, I) \leq {}^2d(E, J) + {}^2d(J, I). \quad [7]$$

Dimostriamo anzitutto la [6]. Indicati con G, H, L tre gruppi di coppie, rispettivamente di punti di E e di I , di E e di J , di J e di I , che diano luogo alle distanze che compaiono nella [6], consideriamo in E un qualunque punto E_h . Esso in G è associato ad un punto di I , che indichiamo con I'_h , e in H è associato ad un punto di J , che potrà essere indicato con J'_h . Sia I''_h il punto che in L è associato ad J'_h .

Se si ha, per ogni h , $I'_h \equiv I''_h$, si deve avere, per la proprietà b) del n. 2,

$$d(E_h, I'_h) \leq d(E_h, J'_h) + d(J'_h, I'_h), \quad [8]$$

e poichè, al variare di h , le coppie (E_h, I'_h) , (E_h, J'_h) , ed (J'_h, I'_h) costituiscono rispettivamente tutte le coppie di G, H e L , sommando la [8], per ogni h , e dividendo ambo i membri per n , si ha la [6].

Se invece non si ha, per ciascun h , $I_h \equiv I''_h$, notiamo che ad ogni E_h corrisponde uno e uno solo I''_h . Si deve avere, sempre per la proprietà b) del n. 2,

$$d(E_h, I''_h) \leq d(E_h, J'_h) + d(J'_h, I''_h). \quad [9]$$

Indicando con G' l'insieme di tutte le coppie (E_h, I''_h) , si ha

$$\frac{1}{n} \sum_{(in G')} d(E_h, I''_h) \leq {}^1d(E, J) + {}^1d(J, I) \quad [10]$$

e, poichè per definizione di distanza è

$${}^1d(E, I) \leq \frac{1}{n} \sum_{(in G')} d(E_h, I''_h), \quad [11]$$

la proprietà [6] è dimostrata completamente.

Dimostriamo ora la [7]. Indichiamo ancora con G, H, L tre gruppi di coppie, rispettivamente di punti E e di punti I , di J e di I , che diano luogo alle distanze che compaiono nella [7]. Fissato E_h in E , siano I'_h e J'_h gli associati di E_h in G ed in H , ed I''_h l'associato di J'_h in L .

Se per ogni h si ha $I'_h \equiv I''_h$ deve essere, per la proprietà b) del n. 2,

$$d(E_h, I'_h) \leq d(E_h, J'_h) + d(J'_h, I'_h) \quad [12]$$

Osserviamo che, poichè da $a_i \leq b_i + c_i$ con $a_i \geq 0$, $b_i \geq 0$, $c_i \geq 0$ ($i = 1, 2, \dots, n$), segue ⁽⁴⁾

$$\sqrt{\sum_1^n a_i^2} \leq \sqrt{\sum_1^n b_i^2} + \sqrt{\sum_1^n c_i^2}, \quad [13]$$

dalla [12] discende

$$\sqrt{\frac{1}{n} \sum_1^n d^2(E_h, I'_h)} \leq \sqrt{\frac{1}{n} \sum_1^n d^2(E_h, J'_h)} + \sqrt{\frac{1}{n} \sum_1^n d^2(J'_h, I'_h)}$$

e perciò la [7]

Se I'_h non coincide con I''_h per ogni h , si ha

$$d(E_h, I''_h) \leq d(E_h, J'_h) + d(J'_h, I''_h), \quad [14]$$

ed, indicato con G l'insieme delle coppie (E_h, I''_h) , si ha per la [13]

$${}^2d(E, I) \leq \sqrt{\frac{1}{n} \sum_{(in G)} d^2(E_h, I''_h)} \leq {}^2d(E, J) + {}^2d(J, I) \quad [15]$$

e in tal modo la [7] è dimostrata completamente.

Le due proprietà $c)$ e $d)$, del n. 2, essendo conseguenza delle due precedenti, valgono anche per le distanze ${}^1d(E, I)$ e ${}^2d(E, I)$ fra insiemi di punti (con i dovuti cambiamenti di parole).

⁽⁴⁾ Infatti da $a_i \leq b_i + c_i$ deriva

$$\begin{aligned} \sum_1^n a_i^2 &\leq \sum_1^n (b_i + c_i)^2 = \sum_1^n (b_i^2 + c_i^2) + 2 \sum_1^n b_i c_i = \\ &= \sum_1^n (b_i^2 + c_i^2) + 2 \sqrt{\sum_1^n b_i^2} \sqrt{\sum_1^n c_i^2} - 2 \left[\sqrt{\sum_1^n b_i^2} \sqrt{\sum_1^n c_i^2} - \sum_1^n b_i c_i \right] = \\ &= \left(\sqrt{\sum_1^n b_i^2} + \sqrt{\sum_1^n c_i^2} \right)^2 - 2 \frac{\sum_1^n b_i^2 \sum_1^n c_i^2 - \left(\sum_1^n b_i c_i \right)^2}{\sqrt{\sum_1^n b_i^2} \sqrt{\sum_1^n c_i^2} + \sum_1^n b_i c_i} = \\ &= \left(\sqrt{\sum_1^n b_i^2} + \sqrt{\sum_1^n c_i^2} \right)^2 - 2 \frac{\sum_{i \neq k} (b_i c_k - b_k c_i)^2}{\sqrt{\sum_1^n b_i^2} \sqrt{\sum_1^n c_i^2} + \sum_1^n b_i c_i} < \left(\sqrt{\sum_1^n b_i^2} + \sqrt{\sum_1^n c_i^2} \right)^2 \end{aligned}$$

e quindi la [13].

Nel caso che la distanza fra un punto qualunque di E ed uno qualunque di I possa essere considerata come *spostamento*, anche le distanze definite al n. 4 possono essere interpretate come il minimo delle medie degli spostamenti che n punti mobili, coincidenti inizialmente con i punti di E , debbono effettuare per portarsi ciascuno su un punto di I .

6. — DISTANZE FRA DUE INSIEMI: CASO GENERALE.

Si è supposto al n. 4 che in ciascun insieme i punti fossero distinti a due a due; tale ipotesi è però inessenziale perchè, se un punto E_h di E ha una molteplicità maggiore di 1, ossia ha una frequenza assoluta v_h , maggiore di 1, si può pensare di sostituire E_h con v_h punti $E'_h, E''_h, \dots, E^{(v_h)}_h$, aventi tutti la stessa distanza da un qualunque punto P di ogni altro insieme. Analogamente se I_k ha una molteplicità n_k si possono sostituire ad esso n_k punti $I'_k, I''_k, \dots, I^{(n_k)}_k$, aventi tutti la stessa distanza da un qualunque punto P di ogni altro insieme. In seguito a tale osservazione, i due insiemi possono considerarsi come costituiti da punti tutti distinti e quindi è possibile estendere immediatamente ad essi quanto esposto ai nn. 4 e 5.

Osserviamo, per gli sviluppi che seguiranno, che, quando alcuni dei punti di E e di I hanno molteplicità maggiore di 1, se si indica con d_{hk} la generica distanza $d(E_h, I_k)$ e con $n^{(j)}_{hk}$ il numero dei punti di E_h che nel generico gruppo $G^{(j)}$ sono associati con altrettanti punti di I_k le [4] e [5] possono scriversi rispettivamente

$${}^1D^{(j)} = \frac{1}{n} \sum_{hk} n^{(j)}_{hk} d_{hk} \quad [16]$$

$${}^2D^{(j)} = \sqrt{\frac{1}{n} \sum_{hk} n^{(j)}_{hk} d_{hk}^2} \quad [17]$$

I minimi delle [16] e delle [17] devono essere determinati sotto le condizioni che gli $n^{(j)}_{hk}$ siano interi non negativi e che

$$\sum_h n^{(j)}_{hk} d_{hk} = n_k \quad \text{e} \quad \sum_k n^{(j)}_{hk} d_{hk} = v_h. \quad [18]$$

È dimostrato che soluzioni delle [18], che diano minimi delle [16] e delle [17], esistono sicuramente ⁽⁵⁾.

(5) GASS, S. I. - *Linear Programming*. Mc. GRAW - HILL 1958, pag. 143.

Se due insiemi E ed I sono costituiti da un diverso numero di punti, che indicheremo rispettivamente con n_E ed n_I , possono ridursi ad averne uno stesso numero moltiplicando la frequenza di ogni punto di E per il quoziente fra un multiplo comune di n_E ed n_I con n_E e moltiplicando la frequenza di ogni punto di I per il quoziente fra lo stesso multiplo comune ed n_I .

Sia n' il minimo comune multiplo di n_E e di n_I , siano E' ed I' gli insiemi di n' elementi ottenuti da E e da I moltiplicando le frequenze dei loro punti rispettivamente per $\frac{n'}{n_E}$ e $\frac{n'}{n_I}$; poniamo $v'_h = n \frac{n'}{n_E} v_h$ e $n'_k = \frac{n'}{n_I} n_k$. Indichiamo con $G^{(i)}$ il generico gruppo che fornisce un'associazione di punti di E' e di I' e relativamente ad esso consideriamo l'espressione [16] cioè la

$${}^1D^{(i)} = \frac{1}{n'} \sum n'^{(i)}_{hk} d_{hk} \quad [19]$$

il cui minimo, sotto le condizioni

$$\sum_h n'^{(i)}_{hk} d_{hk} = n'_k, \quad \sum_k n'^{(i)}_{hk} d_{hk} = v'_h \quad [20]$$

non è altro che ${}^1d(E', I')$. Consideriamo poi il multiplo $n'' = \alpha n'$ di n' , con α intero maggiore di 1, e moltiplichiamo le frequenze di E e di I rispettivamente per $\frac{n''}{n_E}$ e $\frac{n''}{n_I}$, ottenendo i due insiemi E'' ed I'' . Dimostriamo che è ${}^1d(E'', I'') = {}^1d(E', I')$.

Sia $G''^{(i)}$ il generico gruppo che fornisce un'associazione dei punti di E'' e di I'' ; dobbiamo determinare il minimo dell'espressione

$${}^1D''^{(i)} = \frac{1}{n''} \sum n''^{(i)}_{hk} d_{hk} = \frac{1}{n'} \sum \frac{n''^{(i)}_{hk}}{\alpha} d_{hk} \quad [21]$$

sotto le condizioni

$$\sum_h n''^{(i)}_{hk} d_{hk} = \alpha n'_k, \quad \sum_k n''^{(i)}_{hk} d_{hk} = \alpha v'_h, \quad [22]$$

ossia, posto $\bar{n}_{hk} = \frac{n''_{hk}}{\alpha}$, il numero di

$${}^1D''^{(i)} = \frac{1}{n'} \sum \bar{n}_{hk} d_{hk} \quad [23]$$

sotto le condizioni

$$\sum_k \bar{n}_{hk} d_{hk} = n'_k, \quad \sum_k \bar{n}_{hk} d_{hk} = v'_k. \quad [24]$$

Fra le soluzioni del sistema [22] vi sono quelle che rendono minima la [21]. Le radici del sistema [22] si ripartiscono in generale in due classi:

I classe, formata dalle soluzioni costituite, ciascuna, da multipli di α , soluzioni che, come è immediato, divise per α , sono tutte e sole le radici del sistema [20];

II classe, formata da quelle soluzioni nelle quali vi è qualche termine che non è multiplo di α .

Fra le soluzioni che rendono minima la [21] ve ne deve essere almeno una della I classe. Infatti se appartenessero tutte alla II classe, in ognuna di esse vi sarebbe qualche valore \bar{n}_{hk} non intero che renderebbe minima la [23], e ciò è assurdo perchè sussiste un noto teorema ⁽⁶⁾ che afferma che, essendo n'_k e v'_k interi, vi deve essere almeno un sistema di valori, soluzione delle [24], che sia costituito da tutti numeri interi e che renda minima la [23]. Ma allora fra le soluzioni della I classe se ne troverà almeno una che renda minima la [21]; ci si potrà limitare perciò a cercare il minimo, non nel gruppo $G^{(i)}$ ma in $G^{(i)}$ e porre $n''_{hk}^{(i)} = \alpha n'_{hk}^{(i)}$. In tal modo la determinazione del minimo della [21] coincide con quella del minimo della [19].

Si conclude perciò che tutti i multipli comuni n^* di n_E e di n_I danno luogo ad uno stesso valore per il minimo di

$${}^1D^{(i)} = \frac{1}{n} \sum_{hk} n_{hk}^{*(i)} d_{hk} \quad [25]$$

Si dimostra facilmente che lo stesso accade per

$${}^2D^{(i)} = \sqrt{\frac{1}{n} \sum_{hk} n_{hk}^{*(i)} d_{hk}^2} \quad [26]$$

Quanto precede conduce a definire come *distanza del primo ordine* fra E ed I , costituiti da un diverso numero di punti, il

⁽⁶⁾ Vedasi alla pag. 142 dell'opera citata alla nota (5).

minimo delle [25], al variare di j , e come *distanza del secondo ordine* il minimo delle [26]. In conclusione, per definire uno dei due tipi di distanza fra due insiemi, costituiti da un diverso numero di punti, bisogna anzitutto ridurli allo stesso numero di punti e poi usare le definizioni già poste in tale caso.

Quando gli insiemi E ed I sono costituiti dagli stessi punti, quando cioè $E_h \equiv I_h$, ed $n_h = \alpha v_h$, con α intero maggiore di 1, ossia quando può dirsi che I è una « dilatazione » di E , allora è $d(E, I) = 0$. Se, presentandosi questo caso, conveniamo di chiamare ancora *coincidenti* E ed I , le proprietà $a')$ e $b')$ del n. 5 valgono ancora in quanto i due insiemi, considerati nella prima proprietà, ed i tre della seconda si riducono ad avere lo stesso numero di punti ed in tal caso le proprietà $a')$ e $b')$ sono state già dimostrate. D'altra parte come si è già detto, le proprietà $c)$ e $d)$, con i dovuti cambiamenti, sono conseguenza delle proprietà $a')$ e $b')$.

Osserviamo che dalle definizioni poste di distanza fra due insiemi, costituiti da un diverso numero di punti, segue che se, per es., l'insieme E è parte di I , con $n_E < n_I$, ma I non è una « dilatazione » di E , si ha ${}^1d(E, I) > 0$ e ${}^2d(E, I) > 0$ e ciò non deve meravigliare perchè i due tipi di distanza non sono altro che *indici della non coincidenza* fra due insiemi, e i due insiemi in questione non sono coincidenti.

* * *

È importante osservare che ognuno dei due insiemi E ed I può essere interpretato come l'insieme delle determinazioni di una mutabile statistica, per ogni modalità della quale sia definita una distanza da ogni modalità dell'altra mutabile.

7. — CASI PARTICOLARI.

a) Seriazioni.

Se gli insiemi E ed I appartengono ad una stessa retta, è possibile stabilire su essa un sistema di ascisse e usare come distanza fra un punto E_h di ascissa e_h e un punto I_k di ascissa i_k , il numero $|e_h - i_k|$. Attribuendo a ciascuno dei due insiemi il

significato di insieme delle determinazioni di una variabile statistica semplice, le distanze ${}^1d(E, I)$ e ${}^2d(E, I)$, introdotte nei paragrafi 4 e 6 coincidono con i ben noti *indici di dissomiglianza (semplici e quadratici)* fra due distribuzioni statistiche.

b) *Serie territoriali.*

Nel caso delle serie territoriali la distanza fra due modalità può coincidere con la vera distanza fra due punti di due località. La distanza del primo ordine fra le due serie, definita nei paragrafi 4 e 6, non è altro che il minimo delle medie degli spostamenti che debbono effettuare n punti mobili, coincidenti inizialmente con n punti delle località della prima serie, per portarsi su altri n punti delle località della seconda serie.

Il caso delle serie territoriali comprende il *problema dei trasporti* o di Hitchcock-Koopmans, che si può così enunciare: una merce è prodotta (o è in deposito) in alcune località e deve essere trasportata, per il consumo, in un'altra; si conoscono la quantità della merce prodotta (o in deposito) in ogni località e quelle della merce che deve pervenire in ogni località di consumo ed è noto anche il costo del trasporto della quantità unitaria della merce da ogni località di invio a ciascun punto di arrivo. Il problema consiste nello spostare la merce in modo che la spesa di trasporto sia minima.

Per inquadrare il problema dei trasporti nel problema generale trattato, basta misurare la quantità di merce rispetto ad una determinata unità di misura, attribuire alle modalità della serie territoriale, dalle quali parte la merce, una frequenza uguale (o proporzionale) alla quantità di merce in partenza, ed alle modalità della serie territoriale, nelle quali arriva la merce, una quantità uguale (o proporzionale) alla quantità di merce in arrivo e considerare infine come distanza, fra una località da cui parte la merce ed una località in cui deve arrivare, il costo per il trasporto dell'unità di misura della merce dalla prima località all'altra, costo che deve godere della proprietà a cui soddisfa la distanza fra due punti. In tal modo il problema dei trasporti coincide con la ricerca della distanza *del primo ordine* fra due insiemi.

Altri esempi di problemi dei trasporti sono stati trattati (7) e tutti possono essere inquadrati nel caso generale della distanza fra due serie territoriali.

c) *Altre serie sconnesse.*

Esaminiamo il seguente problema: in un'azienda, in cui vi sono tanti operai, tutti ugualmente pagati, quanti posti di lavoro, si conosce il rendimento di ogni operaio in ciascun posto; si debbono distribuire gli operai nei vari posti in modo che sia massimo il rendimento complessivo, ossia che sia minima la somma dei valori assoluti delle differenze fra il rendimento massimo che ogni operaio può dare e quello da lui dato nel posto a cui è assegnato. Tali valori assoluti possono intendersi come distanze operaio-posto e quindi il problema è un caso particolare del problema generale, esposto ai nn. 4 e 6, della ricerca della distanza del primo ordine fra due insiemi.

8. — DETERMINAZIONE DELLE DISTANZE DEL PRIMO E DEL SECONDO ORDINE FRA DUE INSIEMI.

Le definizioni di distanze introdotte rimarrebbero completamente sterili se non esistessero metodi per determinarle, ossia metodi per individuare, nell'insieme di tutte le coppie, due gruppi (che possono coincidere) che forniscono rispettivamente il minimo di ${}^1D^{(j)}$ e ${}^2D^{(j)}$, al variare di j da 1 a $n!$. Osserviamo che il metodo di ricerca è lo stesso perchè per trovare il minimo di ${}^1D^{(j)}$ bisogna rendere minimo $\frac{1}{n} \sum_{hk} n_{hk}^{(j)} d_{hk}$, mentre per determinare il minimo di ${}^2D^{(j)}$ bisogna rendere minimo $\frac{1}{n} \sum_{hk} n_{hk}^{(j)} d_{hk}^2$, ossia in generale si deve trovare il minimo dell'espressione $\frac{1}{n} \sum_{hk} n_{hk}^{(j)} d'_{hk}$ dove d'_{hk} può essere sia d_{hk} che d_{hk}^2 .

a) *Caso delle seriazioni.*

Il primo problema che è stato risolto, nel 1914, è stato quello relativo alle seriazioni: il Gini che introdusse il concetto e la

(7) Si veda, per es. l'opera citata alla nota (5).

misura della *dissomiglianza* ⁽⁸⁾, indicò anche il metodo di calcolo degli indici. Un procedimento rapido per il calcolo degli indici nel caso di seriazioni di frequenza fu introdotto nel 1939 dal Salvemini ⁽⁹⁾ che ideò la *tabella di cograduazione* ⁽¹⁰⁾. È stato dimostrato, infatti, che tale tabella, della quale ho determinato in altro lavoro ⁽¹¹⁾ l'espressione del termine generico, fornisce il minimo per $\sum_{hk} n_{hk}^{(j)} d'_{hk}$. Ciò significa che, se si indicano con

e_1, e_2, \dots, e_r e con i_1, i_2, \dots, i_s le modalità delle due seriazioni E ed I in esame (con $e_1 < e_2 < \dots < e_r$ e $i_1 < i_2 < \dots < i_s$), con $d_{hk} = |e_h - i_k|$ le distanze fra le modalità e_h ed i_k , con v_1, v_2, \dots, v_r le frequenze delle prime, con n_1, n_2, \dots, n_s quelle delle altre (con $\sum_h v_h = \sum_k n_k = n$) e con n_{hk} il termine che occupa la riga h -esima e la colonna k -esima della tabella di cograduazione, avente come margine verticale le v e come margine orizzontale le n , associando n_{hk} delle v_h unità di E_h con altrettante unità di I_k , si ottengono al variare di h e di k , tutte le coppie (E_h, I_k) tali che $\sum_{hk} n_{hk}^{(j)} d'_{hk} = \min$, dove d'_{hk} è uguale a d_{hk} od a d_{hk}^2 .

Perciò la distanza $^1d(E, I)$, ossia l'indice semplice di dissomiglianza fra E ed I , è data da

$$\frac{1}{n} \sum_{hk} n_{hk} |e_h - i_k|; \quad [27]$$

⁽⁸⁾ GINI, C. - *Di una misura della dissomiglianza tra due gruppi di quantità e delle sue applicazioni allo studio delle relazioni statistiche*. Atti del R. Istit. Veneto di Scienze Lettere e Arti. Tomo LXXIV, 1914-15.

GINI, C. - *Nuovi contributi alla teoria delle relazioni statistiche*. Ibidem.

⁽⁹⁾ SALVEMINI, T. - *Sugli indici di omofilia*. Atti della 1^a Riunione della Società Italiana di Statistica, Roma 1939.

⁽¹⁰⁾ Il procedimento per determinare la tabella è denominato nelle pubblicazioni americane «northwest-corner rule». Tale nome è stato coniato negli Stati Uniti dopo 15 anni che in Italia era stata introdotta dal Salvemini la tabella di cograduazione.

⁽¹¹⁾ LETI, G. - *Il termine generico delle tabelle di cograduazione e di contro-graduazione* in appendice a «Note e commenti» - Collana di «Metron», Serie C. Istituto di Statistica della Facoltà di Scienze Statistiche dell'Università di Roma, 1962.

e la distanza ${}^2d(E, I)$, ossia l'indice quadratico didissomiglianza fra E ed I , è data da

$$\sqrt{\frac{1}{n} \sum_{hk} n_{hk} (e_h - i_k)^2}. \quad [28]$$

Per la determinazione di $\sum n_{hk} d'_{hk}$ si potrebbe procedere nel modo seguente: si costruiscono la tabella, ad r righe e ad s colonne $\{d'_{hk}\}$ e la tabella di cograduazione di r righe e di s colonne $\{n_{hk}\}$ e si sommano tutti i prodotti dei termini che in entrambe le tabelle occupano caselle situate nella stessa posizione.

Es.: per calcolare la distanza ${}^2d(E, I)$, fra le due seriazioni

e_h :	1,	3,	5,	6
v_h :	3,	4,	2,	1 ; $n = 10$
i_k :	2,	3,	4,	6
n_k :	2,	2,	4,	2 ; $n = 10$

si costruiscono le due tabelle

$\{d'_{hk}\} =$

$e \backslash i$	2	3	4	6
1	1	4	9	25
3	1	0	1	9
5	9	4	1	1
6	16	9	4	0

$$\{n_{hk}\} =$$

$\begin{array}{c} i \\ e \end{array}$	2	3	4	6	v_k
1	2	1	0	0	3
3	0	1	3	0	4
5	0	0	1	1	2
6	0	0	0	1	1
n_k	2	2	4	2	10

e si calcola la somma

$$1 \cdot 2 + 4 \cdot 1 + 0 \cdot 1 + 1 \cdot 3 + 1 \cdot 1 + 1 \cdot 1 + 0 \cdot 1 = 11$$

La distanza ${}^2d(E, I)$, ossia l'indice quadratico di dissomiglianza misura $\sqrt{\frac{1}{10} {}^2d(E, I)} = \sqrt{1,1}$

b) Caso generale.

Il problema dei trasporti, introdotto da F. L. Hitchcock ⁽¹²⁾ nel 1941, fu formulato come problema di programmazione lineare dal Dantzig ⁽¹³⁾ che ne dette anche una soluzione. I procedimenti seguiti nella risoluzione del problema dei trasporti possono essere usati integralmente per determinare le distanze (del I e del II ordine) dell'insieme E dall'insieme I .

⁽¹²⁾ HITCHCOCK, F. L. - *Distribution of a Product from Several Sources to Numerous Localities*. Journal of Math. Physics. vol. 20, 1941.

⁽¹³⁾ DANTZIG, G. B. - *Application of the Simplex Method to a Transportation Problem*. Cap. XIII del Volume KOOPMANS, T. C., *Activity Analysis of Production and Allocation*. J. Wiley, New York, 1951.

Indicherò brevemente il metodo per la determinazione del minimo di ${}^1D^{(G)}$, ma, per quanto detto all'inizio nel n. 8, il procedimento è il medesimo di quello da usare per la ricerca del minimo di ${}^2D^{(G)}$.

Indicato con E_h ($h = 1, 2, \dots, r$) un punto generico di E , con v_h , la sua frequenza, o molteplicità, e con d_{hk} la sua distanza dal punto I_k ($k = 1, 2, \dots, s$) di I , avente frequenza n_k , si considera la tabella $\{d_{hk}\}$, delle distanze e quella del Salvemini $\{n_{hk}\}$, avente per margine orizzontale le n e come margine verticale le v .

Il problema consiste come si è detto, nella determinazione del minimo della $\sum n_{hk}^{(G)} d_{hk}$, dove $n_{hk}^{(G)}$ indica il numero generico di unità di E_h che sono associate con quelle di I_k in $G^{(G)}$.

L'Herzel ⁽¹⁴⁾ e il Fréchet ⁽¹⁵⁾ hanno dimostrato che, se

$$d_{\alpha\beta} + d_{\gamma\delta} \leq d_{\alpha\delta} + d_{\gamma\beta} \quad [29]$$

per ogni $\alpha < \gamma \leq r$ e $\beta < \delta \leq s$, allora $\frac{1}{n} \sum_{hk} n_{hk} d_{hk}$ è un minimo per $\frac{1}{n} \sum_{hk} n_{hk}^{(G)} d_{hk}$.

È facile vedere ⁽¹⁶⁾ che se $d_{hk} = |e_h - i_k|$, ovvero $d_{hk} = (e_h - i_k)^2$, sono verificate le [29] e perciò, come si è ricordato al n. 6, la tabella del Salvemini è quella che indica l'associazione che fornisce il minimo per ${}^1D^{(G)}$ e ${}^2D^{(G)}$ nel caso delle seriazioni.

In generale se valgono le [29], la determinazione della tabella di cograduazione e il conseguente calcolo di $\frac{1}{n} \sum_{hk} n_{hk} d_{hk}$ esauriscono la ricerca della distanza.

Consideriamo il caso in cui non valgano le [29]. Per determinare una tabella $\{\bar{n}_{hk}\}$ che fornisca il minimo di ${}^1D^{(G)}$, osserviamo anzitutto che nella tabella del Salvemini, $r + s - 1$ caselle, al massimo sono occupate da numeri diversi da zero, men-

⁽¹⁴⁾ HERZEL, A. - *Le tablelle di co- e contrograduazione e la programmazione lineare*, in appendice a « Note e commenti », Collana di Metron - Serie C. Istituto di Statistica della Facoltà di Scienze Statistiche dell'Università di Roma, 1962.

⁽¹⁵⁾ FRECHET, M. - *Les tableaux de corrélation et les programmes linéaires*. « Rev. de l'Institut Intern. de Stat. », vol. 25, n. 1-3, 1957.

⁽¹⁶⁾ Si veda lavoro citato alla nota (14).

tre ognuna delle altre è occupata da zero. È stato dimostrato ⁽¹⁷⁾ che esiste una tabella $\{n_{hk}\}$, che rende minima ${}^1D^{(0)}$, costituita al massimo da $r + s - 1$ numeri positivi, mentre i rimanenti sono tutti nulli. Per determinare tale tabella, Dantzig ⁽¹⁷⁾ ha proposto il seguente procedimento: si consideri la tabella del Salvemini avente per margine le v_h e le n_k , e, per semplicità, supponiamo che sia uguale ad $r + s - 1$ il numero dei termini positivi di detta tabella. Si sostituisca in una qualunque delle caselle occupate da zero un numero positivo, e di conseguenza, poichè le distribuzioni marginali sono fisse, si sottragga lo stesso numero dai due termini, diversi da zero, che occupano, nella tabella del Salvemini considerata, uno la stessa riga e l'altro la stessa colonna e le si sommi ai due termini, diversi da zero, che occupano, nella stessa tabella, rispettivamente uno la stessa colonna e l'altro la stessa riga dei due termini considerati in precedenza. Affinchè ancora il numero dei termini non nulli non superi $r + s - 1$, almeno una delle caselle, in cui si effettua una variazione, dovrà essere occupata, a sostituzione avvenuta da zero, e perciò il numero da sostituire deve essere uguale al più piccolo dei termini da cui va sottratto.

Se ognuna delle nuove tabelle $\{n''_{hk}\}$ a cui si perviene facendo la suddetta sostituzione in una delle possibili caselle, è tale che $\sum_{hk} n''_{hk} d_{hk} \geq \sum_{hk} n_{hk} d_{hk}$, allora $\frac{1}{n} \sum_{hk} n_{hk} d_{hk}$ è la distanza cercata, se, invece fra tutte le tabelle ve ne è qualcuna per la quale $\sum_{hk} n''_{hk} d_{hk} < \sum_{hk} n_{hk} d_{hk}$, si considera quella che dà un massimo per $\sum_{hk} n_{hk} d_{hk} - \sum_{hk} n''_{hk} d_{hk}$, e per quest'ultima si procede di nuovo in modo analogo a quello seguito per la $\{n_{hk}\}$.

Nel seguito ci riferiremo alla tabella $\{n_{hk}\}$, ma ciò che si afferma vale anche per la tabella a cui si perviene successivamente massimizzando le differenze corrispondenti a quella ora considerata. Per abbreviare la ricerca della tabella $\{n''_{hk}\}$ che massimizza la $\sum_{hk} n_{hk} d_{hk} - \sum_{hk} n''_{hk} d_{hk}$, bisogna determinare in quale casella della $\{n_{hk}\}$ va fatta la sostituzione. A questo scopo il Dantzig ha dimostrato che, per ogni termine d_{hk} , avente gli stessi indici di un termine diverso da zero della $\{n_{hk}\}$, si può

⁽¹⁷⁾ Vedasi, per es., l'opera citata alla nota (5).

porre $d_{hh} = u_h + v_h$ e che, posto $\bar{d}_{ij} = u_i + v_j$ per ogni termine avente gli stessi indici di un termine uguale a zero della $\{n_{hk}\}$, se è, per ogni i e j , $\bar{d}_{ij} - d_{ij} \leq 0$, allora la $\{n_{hk}\}$ è la tabella che fornisce il minimo per ${}^1D^{(j)}$. Se invece vi è qualche casella nella quale $\bar{d}_{ij} - d_{ij} > 0$, la $\{n_{hk}\}$ non dà il minimo per la ${}^1D^{(j)}$. La casella nella quale va fatta la sostituzione di cui si è parlato sopra è (come prescrive il metodo generale del simplesso) una di quelle per le quali si ha il massimo delle differenze positive $\bar{d}_{ij} - d_{ij}$.

DISSOMIGLIANZA A PIÙ DIMENSIONI.

I due nuovi tipi di distanza introdotti permettono di generalizzare il concetto di dissomiglianza dal caso di distribuzioni semplici a quello di distribuzioni doppie, triple, ..., n -ple: infatti, date due distribuzioni n -ple E ed I , e definite le distanze fra le modalità dell'una e le modalità dell'altra, le due distanze ${}^1d(E, I)$ e ${}^2d(E, I)$ sono le naturali generalizzazioni dell'indice semplice di dissomiglianza e di quello quadratico fra due distribuzioni semplici.

Anche per chiarire meglio il procedimento esposto nel paragrafo precedente, calcoliamo l'indice semplice di dissomiglianza fra le due seguenti distribuzioni rispetto ai due caratteri X e Y :

Y \ X	X	24	26
	Y		
17		2	3
20		5	—

Y' \ X'	X'	22	24	26
	Y'			
17		1	2	—
20		—	4	3

[30]

La prima tabella individua 3 punti di un insieme E , $E_1 \equiv (17, 24)$, $E_2 \equiv (17, 26)$, $E_3 \equiv (20, 24)$, le cui molteplicità sono rispettivamente 2, 3 e 5. La seconda tabella individua i 4 punti di un insieme I , $I_1 \equiv (17, 22)$, $I_2 \equiv (17, 24)$, $I_3 \equiv (20, 24)$, $I_4 \equiv (20, 26)$, le cui molteplicità sono rispettivamente 1, 2, 4, 3.

Definiamo come distanza fra un punto $E_h \equiv (x_h, y_h)$ e un punto $I_k \equiv (x'_k, y'_k)$ l'espressione $|x_h - x'_k| + |y_h - y'_k|$ (è ovvio

che, mutando la definizione di distanza fra i punti, cambierà di conseguenza il calcolo della distanza fra gli insiemi). Disponiamo le distanze nella tabella

$$\{d_{hk}\} =$$

$E \backslash I$	I_1	I_2	I_3	I_4
E_1	2	0	3	5
E_2	4	2	5	3
E_3	5	3	0	2

La tabella del Salvemini

$E \backslash I$	I_1	I_2	I_3	I_4	Totale
E_1	1	1			2
E_2		1	2		3
E_3			2	3	5
Totale	1	2	4	3	10

[31]

dà luogo al valore ${}^1D^{(t)} = \frac{20}{10} = 2$.

Si debbono perciò ricercare i numeri u_h e v_k tali che

$$u_1 + v_1 = d_{11} = 2$$

$$u_1 + v_2 = d_{12} = 0$$

$$u_2 + v_2 = d_{22} = 2$$

$$u_2 + v_3 = d_{23} = 5$$

$$u_3 + v_3 = d_{33} = 0$$

$$u_3 + v_4 = d_{34} = 2$$

Tale sistema di equazioni è costituito da 6 equazioni in 7 incognite e perciò ad una incognita deve essere dato ad arbitrio un valore: attribuiamo per es., a v_1 il valore 2.

Un metodo semplice per risolvere il sistema è quello di costruire il quadro

$v_k \backslash u_h$				
	2	0		
		2	5	
			0	2

e di calcolarne i due margini, dopo aver posto, $v_1 = 2$

$v_k \backslash u_h$	2	0	3	5
0	2	0		
2		2	5	
-3			0	2

in modo che sia $u_h + v_h = d_{hh}$. Si calcolano quindi i termini delle caselle vuote ponendo $\bar{d}_{ij} = u_i + v_j$, in modo da pervenire alla tabella

$\{\bar{d}_{ij}\} =$

$\begin{array}{c} v_j \\ \hline u_i \end{array}$	2	0	3	5
0	2	0	3	5
2	4	2	5	7
-3	-1	-3	0	2

Vi è una casella nella quale si ha $\bar{d}_{ij} - d_{ij} > 0$ e precisamente quella che si trova sulla seconda riga e la quarta colonna, perciò la tabella del Salvemini non fornisce il minimo per la distanza dei due insiemi e in tale casella della [31] si deve sostituire allo zero un numero positivo θ . Si scrive perciò

1	1		
	1	$2 - \theta$	θ
		$2 + \theta$	$3 - \theta$

da cui si ha, per eliminare un termine, $\theta = 2$ e quindi la tabella

$\begin{array}{c} \diagdown \\ E \end{array}$	I	I_1	I_2	I_3	I_4	Totale
E_1	I	I				2
E_2			I		2	3
E_3				4	I	5
Totale	I	2	4	3		10

che fornisce per ${}^1D^{(i)}$ il valore $\frac{12}{10} = 1,2$.

Tale valore è il minimo possibile perchè gli u_i e v_j e la tabella dei \bar{d}_{ij} , relativi all'ultima tabella, dati da

$\begin{array}{c} \diagdown \\ v_j \\ u_i \end{array}$	2	0	— I	I
0	2	0	— I	I
2	4	2	1	3
I	3	I	0	2

sono tali che ogni $\bar{d}_{ij} - d_{ij}$ è non superiore a zero. Perciò la distanza fra i due insiemi è 1,2 e noi assumiamo questo numero come *indice semplice di dissomiglianza* delle due distribuzioni doppie [30]

Ricordiamo che il Salvemini ⁽¹⁸⁾ ha dato un'estensione al caso di distribuzioni doppie dell'indice semplice di dissomiglianza fra distribuzione semplici, generalizzando la proprietà che l'indice semplice di dissomiglianza è il rapporto, rispetto al numero dei casi, dell'area compresa fra le due curve cumulative relative alle due variabili semplici. Il Salvemini ha costruito le due superficie cumulative relative alle due distribuzioni doppie e ha definito come indice semplice di dissomiglianza il rapporto fra il volume compreso fra le due superficie e il numero n dei casi.

Dobbiamo notare che la generalizzazione dell'indice semplice di dissomiglianza proposta in questo lavoro è diversa da quella data dal Salvemini. Infatti il valore dell'indice del Salvemini non varia se si traslano gli assi coordinati, ma in generale cambia se questi vengono ribaltati, in quanto ad ogni ribaltamento corrisponde generalmente, come ha mostrato il Gini ⁽¹⁹⁾, una diversa superficie cumulativa, anche se si mutano di conseguenza le disuguaglianze che compaiono nella definizione di tale superficie. Invece l'indice di dissomiglianza, ottenuto mediante le distanze, è un invariante rispetto alle traslazioni ed ai ribaltamenti degli assi, in quanto tali trasformazioni di coordinate non comportano una variazione della distanza $|x_h - x'| + |y_h - y'_k|$, (od anche della distanza $\sqrt{(x_h - x'_k)^2 + (y_h - y'_k)^2}$), perchè l'aggiunta di una stessa quantità a tutte le x o a tutte le y o il cambiamento di segno di tutte le x , o di tutte le y , o di tutte le x e y , non fa variare il valore di tali espressioni.

In particolare, la differenza fra le due generalizzazioni risulta anche dal fatto che, nell'esempio dato, l'indice semplice di dissomiglianza, ottenuto mediante le distanze, vale 1,2, in base alle distanze definite fra i punti, mentre il valore di quello introdotto dal Salvemini, per le stesse distribuzioni e nello stesso sistema di assi coordinati è diverso.

⁽¹⁸⁾ SALVEMINI, T. - *Dissomiglianza a più dimensioni*. Atti della IX, X e XI Riunione della Società Italiana di Statistica, 1951.

⁽¹⁹⁾ GINI, C. - *Sulla dissomiglianza e le relazioni statistiche* (in corso di sistemazione). Tale lavoro riassume, integra ed espone in maniera sistematica i contributi su tale argomento del Gini e della sua Scuola.

10. — INDICE SEMPLICE DI DISSOMIGLIANZA TRA LE DISTRIBUZIONI DEI MATRIMONI SECONDO L'ETÀ DEGLI SPOSI IN DUE REGIONI ITALIANE.

Per calcolare l'indice semplice di dissomiglianza fra due distribuzioni doppie effettive, ho preso in considerazione le distribuzioni dei matrimoni celebrati nel 1958 in Calabria e in Piemonte ⁽²⁰⁾, secondo classi d'età ⁽²¹⁾ degli sposi. Le distribuzioni sono riportate nelle tabelle 1 e 2.

TAB. 1. — *Matrimoni celebrati in Calabria nel 1958, secondo classi d'età degli sposi.*

ETÀ DEGLI SPOSI (in anni)	ETÀ DELLE SPOSE (in anni)								Totale
	Fino a 15 (14,0) x_1	15-21 (18,0) x_2	21-25 (23,0) x_3	25-30 (27,5) x_4	30-35 (32,5) x_5	35-40 (37,5) x_6	40-50 (45,0) x_7	Oltre 50 (55,0) x_8	
Fino a 18 - (17,0) - y_1	5	69	20	6	—	—	—	—	100
18-21 - (19,5) - y_2	31	805	335	84	13	4	—	—	1.272
21-25 - (23,0) - y_3	30	2.249	1.782	490	64	9	1	1	4.626
25-30 - (27,5) - y_4	8	2.052	2.409	1.250	182	31	6	—	5.938
30-35 - (32,5) - y_5	4	491	854	787	274	38	4	—	2.452
35-40 - (37,5) - y_6	—	63	173	312	171	58	19	—	796
40-50 - (45,0) - y_7	—	15	42	137	151	115	71	3	534
50-60 - (55,0) - y_8	—	2	3	17	37	40	77	41	217
Oltre 60 - (65,0) - y_9	—	1	6	18	25	18	61	97	226
TOTALE . . .	78	5.747	5.624	3.101	917	313	239	142	16.161

Il numero delle classi della distribuzione del Piemonte è 55, mentre quello della Calabria è 59. A causa dell'alto numero di classi è stato impossibile calcolare l'indice semplice di dissomiglianza senza far ricorso ad un elaboratore elettronico dei dati ⁽²²⁾.

Per procedere nel calcolo è necessario ridurre le due distribuzioni delle tabelle 1 e 2 ad avere lo stesso numero di ele-

⁽²⁰⁾ Fonte: *Annuario di statistiche demografiche*. 1958. Istituto Centrale di Statistica. Roma, 1961, pag. 38-39.

⁽²¹⁾ Per la definizione delle classi d'età vedasi a pag. 9 del citato annuario.

⁽²²⁾ Il calcolo di una tabella $\{n_{hk}\}^t$, che fornisce il minimo, è stato gentilmente eseguito presso il Centro di calcolo scientifico di Roma dell'I.B.M. - Italia.

TAB. 2. — *Matrimoni celebrati in Piemonte nel 1958, secondo classi d'età degli sposi.*

ETÀ DEGLI SPOSI (in anni)	ETÀ DELLE SPOSE (in anni)								Totale
	Fino a 15 (14,0) x'_1	15-21 (18,0) x'_2	21-25 (23,0) x'_3	25-30 (27,5) x'_4	30-35 (32,5) x'_5	35-40 (37,5) x'_6	40-50 (45,0) x'_7	Oltre 50 (55,0) x'_8	
Fino a 18 - (17,0) - y'_1	—	9	2	—	—	—	—	—	11
18-21 - (19,5) - y'_2	4	369	98	25	2	1	—	—	499
21-25 - (23,0) - y'_3	8	1.891	2.607	627	63	7	—	—	5.203
25-30 - (27,5) - y'_4	—	2.322	5.127	3.226	422	61	8	1	11.167
30-35 - (32,5) - y'_5	—	461	1.479	1.869	756	157	27	2	4.751
35-40 - (37,5) - y'_6	—	59	265	538	479	183	32	4	1.560
40-50 - (45,0) - y'_7	1	12	38	183	321	270	228	42	1.095
50-60 - (55,0) - y'_8	—	3	3	13	34	75	195	143	466
Oltre 60 - (65,0) - y'_9	—	1	—	—	3	9	58	223	294
TOTALE . . .	13	5.127	9.619	6.481	2.080	763	548	415	25.046

menti. Poichè il minimo comune multiplo di 16.161 e 25.046 è il loro prodotto 404.768.408, si sono portate le due distribuzioni ad avere come totale tale numero moltiplicando i dati relativi alla Calabria per 25.046 e quelli relativi al Piemonte per 16.161.

I risultati di queste operazioni sono numeri aventi al massimo 8 cifre (tranne il totale che ne ha nove); ma poichè il procedimento elettronico di calcolo del problema di Hitchcock, predisposto per l'elaboratore che ha eseguito il calcolo, prevede che tali numeri (eccetto il totale) siano al massimo di 5 cifre, è stato necessario dividere ogni numero per 1000, arrotondando l'ultima cifra rimasta. Le due distribuzioni, che sono state quindi ridotte ad avere entrambe $n = 404.768$ unità, sono riportate nelle tabelle 3 e 4.

Per il calcolo delle distanze si è sostituita ad ogni classe la semisomma degli estremi, tranne che per le classi estreme che sono state sostituite da valori ricavati in base alla distribuzione in Italia degli sposi per singole età, distribuzione che è nota. Indicata allora con x_h l'età rappresentativa dell' h -esima classe d'età delle spose della Calabria, con x'_k quella della k -esima classe d'età delle spose del Piemonte, e con y_i e y'_j le età rappresentative rispettivamente dell' i -esima classe degli sposi della Calabria e dell' j -esima classe degli sposi del Piemonte, si è defi-

TAB. 3. — *Distribuzione di 404.768 matrimoni proporzionale alla distribuzione dei matrimoni celebrati in Calabria.*

$\begin{array}{c} X \\ \backslash \\ Y \end{array}$	14,0 x_1	18,0 x_2	23,0 x_3	27,5 x_4	32,5 x_5	37,5 x_6	45,0 x_7	55,0 x_8
17,0 - y_1	125	1.728	501	150	—	—	—	—
19,5 - y_2	776	20.162	8.390	2.104	326	100	—	—
23,0 - y_3	751	56.329	44.632	12.273	1.603	225	25	25
27,5 - y_4	200	51.394	60.336	31.308	4.558	776	150	—
32,5 - y_5	100	12.298	21.389	19.711	6.863	952	100	—
37,5 - y_6	—	1.578	4.333	7.814	4.283	1.453	476	—
45,0 - y_7	—	376	1.052	3.431	3.782	2.880	1.778	75
55,0 - y_8	—	50	75	426	927	1.002	1.929	1.027
65,0 - y_9	—	25	150	451	626	451	1.528	2.430

nita come distanza fra la coppia di età x_h, y_i e quella d'età x'_h e y'_i l'espressione

$$|x_h - x'_h| + |y_i - y'_i| \quad [32]$$

I valori delle distanze ottenute sono stati sistemati in una tabella nella quale sono state considerate solo le distanze fra coppie d'età aventi una frequenza diversa da zero. In tale tabella le coppie d'età sono state ordinate, indipendentemente per la Calabria e per il Piemonte, considerando come prima la coppia costituita dalle età più basse dello sposo e della sposa e

TAB. 4. — *Distribuzione di 404.768 matrimoni proporzionale alla distribuzione dei matrimoni celebrati in Piemonte.*

$\begin{array}{c} X' \\ \backslash \\ Y' \end{array}$	14,0 x'_1	18,0 x'_2	23,0 x'_3	27,5 x'_4	32,5 x'_5	37,5 x'_6	45,0 x'_7	55,0 x'_8
17,0 - y'_1	—	145	32	—	—	—	—	—
19,5 - y'_2	65	5.963	1.584	404	32	16	—	—
23,0 - y'_3	129	30.560	42.132	10.133	1.018	113	—	—
27,5 - y'_4	—	37.526	82.857	52.135	6.820	986	129	16
32,5 - y'_5	—	7.450	23.902	30.205	12.218	2.537	436	32
37,5 - y'_6	—	954	4.283	8.695	7.741	2.958	517	65
45,0 - y'_7	16	194	614	2.958	5.188	4.363	3.685	679
55,0 - y'_8	—	49	49	210	550	1.212	3.151	2.311
65,0 - y'_9	—	16	—	—	49	145	937	3.604

come successiva ad ogni coppia o quella costituita dalla medesima età dello sposo e dall'età della sposa immediatamente più grande o, esaurite le coppie aventi una medesima età dello sposo, la coppia la cui età dello sposo è quella immediatamente superiore a quella precedentemente considerata e l'età della

TAB. 5. — Valori, diversi da zero, di una tabella $\{\bar{n}_{hk}\}$ che fornisce l'indice di dissomiglianza e distanze relative a tali valori.

h	k	\bar{n}_{hk}	d_{hk}^*	h	k	\bar{n}_{hk}	d_{hk}^*	h	k	\bar{n}_{hk}	d_{hk}^*
I	3	65	2,5	19	16	200	9,0	40	38	432	—
I	14	4	29,5	20	16	51.394	5,0	40	39	620	4,5
I	35	40	61,5	21	16	24.711	—	41	39	759	—
I	36	16	28,0	21	17	35.625	4,5	41	40	2.672	5,0
2	1	145	—	22	17	4.237	—	42	41	3.415	5,0
2	4	1.252	2,5	22	18	5.941	5,0	42	47	367	10,0
2	8	16	22,0	22	24	21.130	5,0	43	42	2.880	7,5
2	14	9	25,5	23	19	986	5,0	44	42	580	—
2	20	104	37,5	23	25	3.572	5,0	44	43	679	10,0
2	21	16	47,5	24	26	776	5,0	44	49	519	10,0
2	27	186	42,5	25	27	150	5,0	45	50	75	10,0
3	2	32	—	26	22	100	4,0	46	44	49	—
3	5	469	2,5	27	22	7.350	—	46	47	1	14,5
4	6	150	2,5	27	23	4.948	5,0	47	45	49	—
5	4	647	4,0	28	23	18.954	—	47	47	26	9,5
5	9	129	3,5	28	24	2.435	4,5	48	46	210	—
6	4	4.064	—	29	24	6.640	—	48	47	156	5,0
6	10	16.098	3,5	29	25	8.646	5,0	48	48	60	10,0
7	5	1.115	—	29	31	4.425	5,0	49	48	927	5,0
7	11	7.275	3,5	30	26	1.761	5,0	50	49	1.002	7,5
8	6	254	—	30	32	5.102	5,0	51	49	1.630	—
8	12	1.850	3,5	31	33	952	5,0	51	50	299	10,0
9	7	32	—	32	27	100	—	52	50	1.027	—
9	13	294	3,5	33	29	954	—	53	51	16	—
10	14	100	3,5	33	30	624	5,0	53	52	9	14,5
11	10	751	4,0	34	30	3.659	—	54	52	40	9,5
12	10	13.711	—	34	31	674	4,5	54	53	110	14,5
12	11	5.092	5,0	35	31	3.596	—	55	53	35	10,0
12	15	37.526	4,5	35	32	2.639	5,0	55	54	416	17,5
13	11	29.765	—	35	39	1.579	7,5	56	54	521	12,5
13	12	8.283	4,5	36	33	1.767	5,0	56	55	105	22,5
13	16	6.552	4,5	36	40	2.516	7,5	57	55	451	17,5
13	28	32	41,0	37	33	239	—	58	55	1.528	10,0
14	17	12.273	4,5	37	34	266	7,5	59	50	910	10,0
15	13	724	—	37	41	948	7,5	59	55	1.520	—
15	18	879	4,5	38	34	251	—				
16	48	225	31,0	38	42	225	7,5				
17	20	25	4,5	39	37	194	—				
18	35	25	14,5	39	38	182	5,0			404768	

sposa è la più bassa. Non si è ritenuto opportuno riportare la tabella delle distanze perchè di notevoli dimensioni e perchè facilmente ricavabile mediante la [32].

Essendo stato dato lo stesso ordinamento ai due insiemi di coppie, i valori piccoli delle distanze si dispongono generalmente nei pressi delle caselle occupate da zero, ossia di quelle caselle nelle quali le coppie corrispondenti sono costituite sia da femmine che da maschi aventi rispettivamente la stessa età.

Con lo stesso ordinamento delle coppie della tabella delle distanze si sono ottenuti dalla tab. 3 i valori v_h (frequenza della coppia della Calabria, che nell'ordinamento dato occupa l' h -esimo posto, $h = 1, 2, \dots, 59$) e dalla tab. 4 i valori n_k (frequenza della coppia del Piemonte che nell'ordinamento dato occupa il k -esimo posto, $k = 1, 2, \dots, 55$).

Il calcolatore elettronico ha fornito una delle soluzioni possibili del problema, ossia una tabella $\{\bar{n}_{hk}\}$ che fornisce il minimo di ${}^1D^{(i)}$. Nella tab. 5 sono stati riportati i valori di tale tabella che sono diversi da zero e che sono in numero di $55 + 59 - 1 = 113$, e, accanto ad ognuno, il valore della distanza d_{hk} fra la h -esima coppia della Calabria e quella k -esima del Piemonte.

Dalla tab. 5 si osserva che vi sono molti valori \bar{n}_{hk} (diversi da zero) corrispondenti a distanze nulle, ma vi sono anche molte distanze zero a cui non corrisponde alcun valore, diverso da zero, di \bar{n}_{hk} (almeno per la soluzione ottenuta). Però tutti i 113 termini \bar{n}_{hk} diversi da zero, tranne una decina, sono disposti nei pressi delle caselle corrispondenti a distanze nulle, ossia con gli indici h e k differenti di poco fra loro, e ciò era da attendersi in quanto, come si è detto, tali caselle sono generalmente occupate da valori piccoli delle distanze.

L'indice di dissomiglianza è dato da

$$\frac{\sum_{hk} \bar{n}_{hk} d_{hk}^*}{n} = \frac{1.377.628,5}{404.768} = 3,404 \text{ anni.}$$

Ciò significa che ogni coppia di una delle due distribuzioni deve subire in media uno spostamento di 3,4 anni, affinché la distribuzione di cui fa parte si porti a coincidere con l'altra distri-

buzione, intendendosi per spostamenti la somma dei valori assoluti della differenza delle età delle spose e della differenza delle età degli sposi.

Per poter dare una chiara valutazione del risultato sarebbe necessario determinare lo stesso indice fra distribuzioni di matrimoni in altre regioni e poi procedere ai confronti dei risultati ottenuti. Si è voluto però fare questa applicazione per mostrare che è possibile per distribuzioni effettive ottenere il valore dell'indice di dissomiglianza anche se si è dovuti ricorrere, per ottenerlo, all'ausilio di un calcolatore elettronico. Lo studioso di problemi teorici di statistica non può oggi fare a meno di questi calcolatori, e, nell'elaborazione dei suoi metodi, non solo non ne deve ignorare l'esistenza, ma deve usufruire di essi il più possibile.

A note on Gini's mean difference

A. R. KAMAT

1) *Introduction.* — Gini's (1912) mean difference as a measure of dispersion is of considerable theoretical and historical interest. If x_i ($i = 1, 2, \dots, n$) is a random sample of observations, then Gini's mean difference for the sample may be defined by

$$g = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=i+1}^n |x_i - x_j| \quad [1]$$

Nair (1936) and Lomnicki (1952) gave expressions for the variance of g for any parental distribution. The present author (1953) derived the third moment of g , when the parental distribution is normal $N(\mu, \sigma)$, and obtained an approximate expression for the fourth moment (1960.*). Nabeya (1961) has recently given an exact expression for the raw fourth moment, $\mu'_4(g)$ for the normal parent population and it is now possible to calculate exactly the fourth central moment and the β_2 constant of g . It is proposed to put these and other relevant results on record in this note.

2) *The expression for $\mu_4(g)$.* — The expression for $\mu'_4(g)$ given by Nabeya (1961) on p. 83 can be further simplified by mak-

* RAMSUBBAN (1956) has also independently derived an approximate expression for the fourth moment. For consideration of exact dates it may be mentioned that although the volume in which Ramsubban's result appears is dated 1956, it actually came out in 1958. My results which were in an unpublished form since 1955 were submitted for publication in the Commemoration Volume in honour of Professor GINI in 1958. A comparison of the numerical accuracy of the two approximations is given in section 3.

ing use of the equalities

$$\begin{aligned}\pi/2 &= \sin^{-1}(1/3) + 2 \sin^{-1}(1/\sqrt{3}) = \sin^{-1}(2/3) + 2 \sin^{-1}(1/\sqrt{6}) \\ &= \sin^{-1}(1/4) + 2 \sin^{-1}(\sqrt{3}/8)\end{aligned}$$

and can be rewritten as

$$\begin{aligned}\mu'_4(g) &= \frac{16 \sigma^4}{n^3 (n-1)^3 \pi^2} [6 \pi^2 + (72 \sqrt{3} \pi + 26 \pi^2) (n-2) + \\ &+ (112 \pi + 144 \sqrt{3} \pi + 21 \pi^2) (n-2)^{[2]} + \\ &+ (100 \sqrt{5} + 48 \sqrt{3} \pi + 120 \pi + 5 \pi^2) (n-2)^{[3]} + \\ &+ (36 + 64 \sqrt{2} + 30 \pi + 4 \sqrt{3} \pi + (1/3) \pi^2) (n-2)^{[4]} + \\ &+ (12 \sqrt{3} + 2 \pi) (n-2)^{[5]} + (n-2)^{[6]}] \quad [2]\end{aligned}$$

The expression for $\mu_4(g)$ is then given by

$$\mu_4(g) = \frac{16 \sigma^4}{n^3 (n-1)^3 \pi^2} [A_4 n^4 + A_3 n^3 + A_2 n^2 + A_1 n + A_0] \quad [3]$$

where

$$\begin{aligned}A_4 &= (1/3) \pi^2 + 4 \sqrt{3} \pi - 8 \pi - 48 \sqrt{3} + 84, \\ A_3 &= 12 \pi - 8 \sqrt{3} \pi + (1/3) \pi^2 - 512 \sqrt{2} + 1080 \sqrt{3} + 100 \sqrt{5} - 1368, \\ A_2 &= 8 \pi - 4 \sqrt{3} \pi - (1/3) \pi^2 + 3840 \sqrt{2} - 5784 \sqrt{3} - 900 \sqrt{5} + 6600, \\ A_1 &= 20 \pi - 16 \sqrt{3} \pi - (1/3) \pi^2 - 9472 \sqrt{2} + 11952 \sqrt{3} + 2600 \sqrt{5} - 13092, \\ A_0 &= 48 \sqrt{3} \pi - 48 \pi + 7680 \sqrt{2} - 8640 \sqrt{3} - 2400 \sqrt{5} + 9360\end{aligned}$$

Evaluating the coefficients numerically, we have

$$\begin{aligned}\mu_4(g) &= n^{-3} (n-1)^{-3} (1.271428 n^4 - 0.645017 n^3 + \\ &+ 0.023198 n^2 + 0.481067 n + 0.111367) \sigma^4 \quad [4]\end{aligned}$$

For the sake of completeness the first three central moments of g are reproduced below:

$$\mu_1(g) = \frac{2\sigma}{\sqrt{\pi}} = (1.128379)\sigma$$

$$\begin{aligned}\mu_2(g) &= \frac{4\sigma^2}{n(n-1)\pi} [(2\sqrt{3}-4 + (1/3)\pi)n + (6-4\sqrt{3} + (1/3)\pi)] \\ &= n^{-1}(n-1)^{-1}(0.651006n + 0.151508)\sigma^2\end{aligned}\quad [5]$$

$$\begin{aligned}\mu_3(g) &= \frac{8\sigma^3}{n^2(n-1)^2\pi^{3/2}} [(16\sqrt{2} + 40 - 36\sqrt{3})n^2 + \\ &+ (144\sqrt{3} - 136 - 80\sqrt{2})n + (96\sqrt{2} + 120 - 144\sqrt{3} - 2\pi)] \\ &= n^{-2}(n-1)^{-2}(0.393063n^2 + 0.399734n + 0.094823)\sigma^3\end{aligned}$$

(Working with ten figure accuracy for the irrationals involved it was found that the coefficient of n^4 in the numerical expression of the fourth moment and the coefficient of n and the constant term in the numerical expression of the third moment given by us previously (1953 and 1960) need correction in the last digit. This has been done in [4] and [5] above).

3) *Tabulation of constants.* — With the derivation of the exact value of the fourth moment the distribution theory of g (for normal parental population) may be considered to have completed a definite stage when it is appropriate to put on record the moment constants and the β_1 , β_2 constants of g . (We have not seen any tables of these constants published so far.) The table below (Table 1) gives μ_2 , μ_3 , μ_4 , and σ , β_1 , β_2 for the sample sizes $n = 2$ (1) 20, 25 (5) 50, 60 (10) 100. The constant β_2 could be calculated in two ways: (i) by straightforward evaluation of $\mu_4/3\mu_2^2$; (ii) by expressing $\mu_4 - 3\mu_2^2 = c_3n^3 + c_2n^2 + c_1n + c_0$, and then writing β_2 as

$$\beta_2 = 3 + \frac{(0.021352n^3 + 0.336881n^2 + 0.339225n + 0.068697)}{n(n-1)(0.261427n^2 + 0.121683n + 0.014160)} [6]$$

It is clear from [6] that it will achieve better numerical accuracy than method (i). We may state here that all the values in Table 1 are expected to be correct to the last digit as the calcu-

TABLE I. — *Moment Constants of Gini's Mean Difference (g/σ) for normal population. Mean = 1.128379*

n	μ_2	μ_3	μ_4	σ	β_1	β_2
2	0.726760	0.616636	2.043625	0.8525	0.9906	3.8692
3	0.350755	0.134211	0.404322	0.5922	0.4174	3.2864
4	0.229628	0.055436	0.165863	0.4792	0.2538	3.1456
5	0.170327	0.029800	0.089639	0.4127	0.1797	3.0898
6	0.135252	0.018493	0.056010	0.3678	0.1382	3.0618
7	0.112108	0.012558	0.038280	0.3348	0.1119	3.0457
8	0.095706	0.009072	0.027805	0.3094	0.0939	3.0356
9	0.083480	0.006854	0.021106	0.2889	0.0807	3.0287
10	0.074017	0.005358	0.016566	0.2721	0.0708	3.0237
11	0.066478	0.004302	0.013347	0.2598	0.0630	3.0201
12	0.060330	0.003529	0.010982	0.2456	0.0567	3.0173
13	0.055222	0.002947	0.009195	0.2350	0.0515	3.0152
14	0.050910	0.002498	0.007810	0.2256	0.0473	3.0134
15	0.047222	0.002144	0.006717	0.2173	0.0436	3.0120
16	0.044032	0.001860	0.005837	0.2098	0.0405	3.0108
17	0.041245	0.001629	0.005120	0.2031	0.0378	3.0099
18	0.038790	0.001438	0.004527	0.1970	0.0354	3.0090
19	0.036610	0.001279	0.004032	0.1913	0.0333	3.0083
20	0.034662	0.001145	0.003614	0.1862	0.0315	3.0077
25	0.027378	0.000710	0.002253	0.1655	0.0246	3.0055
30	0.022623	0.000483	0.001538	0.1504	0.0202	3.0043
35	0.019275	0.000350	0.001116	0.1388	0.0171	3.0035
40	0.016790	0.000265	0.000846	0.1296	0.0148	3.0029
45	0.014872	0.000208	0.000664	0.1220	0.0131	3.0025
50	0.013348	0.000167	0.000535	0.1155	0.0117	3.0022
60	0.011077	0.000115	0.000368	0.1052	0.0097	3.0017
70	0.009466	0.000084	0.000269	0.0973	0.0083	3.0014
80	0.008265	0.000064	0.000205	0.0909	0.0072	3.0012
90	0.007334	0.000050	0.000161	0.0856	0.0064	3.0011
100	0.006591	0.000041	0.000130	0.0812	0.0057	3.0010

lations were done with eight-figure accuracy throughout for the evaluation of moments.

It is interesting to compare the exact expression for $\mu_4(g)$ given in [4] with the approximate expressions for the same obtained by us (1960) and Ramsubbhan (1956) respectively, by using power series of correlation coefficients for evaluating certain expectations appearing in the fourth moment. This comparison as well as the comparison of numerical values of μ_4 and β_2 shows

that our approximate results happened to be nearer the correct values than those given by Ramsubban.

4) *Approximating the distribution of g .* — It is known that g is asymptotically normally distributed. (See: Lomnicki and Zaremba, 1957). From the values of β_1 , β_2 given above it seems reasonable to assume that the normal approximation should be adequate for $n > 30$ and perhaps even for n as low as 20. Considering small samples, we know that for $n = 3$, $g = (2/3)w$, where w is the sample range. Hartley and Pearson (1951) have shown that for small samples the χ -distribution gives a good approximation for the range. The present author (1953) had therefore suggested the same approximation for small samples. If we consider the points on the β_1 , β_2 diagram (for Pearson Curves) it is found that they lie in the Pearson Type I region, very close to the points of the χ -distribution. The following table compares the β_1 , β_2 constants of g , with those of the χ -distribution, for the same values of the coefficient of variation $\sqrt{\mu_2}/\mu_1$. The latter have been obtained by interpolation from Pearson and Hartley (1956). The comparison appears to confirm our suggestion of the χ -approximation.

n	$\sqrt{\mu_2}/\mu_1$	Values for g		Values for χ -distribution	
		β_1	β_2	β_1	β_2
5	0.3657	0.1797	3.0898	0.1672	3.0615
10	0.2411	0.0708	3.0237	0.0648	3.0111
15	0.1926	0.0436	3.0120	0.0398	3.0044
20	0.1650	0.0315	3.0077	0.0286	3.0023

I am thankful to Mr. R. S. Rao for help in calculations of values in Table 1.

REFERENCES

- GINI, C. (1912): *Variabilità e Mutabilità*. « Studi Economico-Giuridici della R. Università di Cagliari », III year, part. 2. Reproduced in *Variabilità e Concentrazione*, 1^a Edizione, Giuffrè, Milano 1939, 2^a Edizione, Veschi, Roma 1955.
- HARTLEY, H. O. and PEARSON, E. S. (1951): *Moment constants for the distribution of range in normal samples*. « Biometrika », 38, 463.
- KAMAT, A. R. (1953): *The third moment of Gini's mean difference*. « Biometrika », 40, 451.
- KAMAT, A. R. (1960): *Contributions to the theory of Gini's mean difference*. « Commemoration Volume in Honour of Professor Gini ».
- LOMNICKI, Z. A. (1952): *The Standard Error of Gini's Mean Difference*. « Annals of Mathematical Statistics », 23, 635.
- LOMNICKI, Z. A. and ZAREMBA, S. K. (1957): *A Note on Central Limit Theorem*. « Mathematische Zeitschrift », 66, 490.
- NABEYA, S. (1961): *Absolute and incomplete moments of the multivariate normal distribution*. « Biometrika », 48, 77.
- NAIR, U. S. (1936): *A Note on Gini's Mean Difference*. « Biometrika », 28, 428.
- PEARSON, E. S. and Hartley, H. O. (1956): « *Biometrika tables for Statisticians* », Vol. I, Cambridge U.P.
- RAMSUBBAN, T. A. (1956): *A χ -approximation to Gini's mean difference*. « Journal of the Indian Society of Agricultural Statistics », 8, 116.

GERALD J. GLASSER

**Relationships between the mean difference
and other measures of variation**

1. — INTRODUCTION

The coefficient of mean difference, introduced by C. Gini, is defined as

$$\Delta = \frac{1}{N(N-1)} \sum_{i \neq j} |x_i - x_j| \quad [1]$$

which suggests one simple interpretation for the parameter as measuring the average difference between pairs of observations. Viewed in this way, this measure of dispersion is a simple descriptive parameter suitable for many problems.

The coefficient may, however, be written in several alternative forms. Thus, suppose that the observations are considered arrayed in order of magnitude (observations of equal value may be placed in any order) so that i is the rank of $X_{(i)}$. Then, as is easily verified, Δ may be written as

$$\Delta = \frac{2}{N(N-1)} \sum_{i=1}^N (2i - N - 1) X_{(i)} \quad [2]$$

or where μ is the arithmetic mean of the observations, as

$$\Delta = \frac{2}{N(N-1)} \sum_{i=1}^N 2i (X_{(i)} - \mu) \quad [3]$$

or

$$\Delta = \frac{2}{N(N-1)} \sum_{i=1}^N (2i - N - 1) (X_{(i)} - \mu) \quad [4]$$

Thus, in addition to providing relatively simple means of computation these formulas show that Δ may also be interpreted as a weighted average of the deviations of the observations from

their mean. In this sense, it is similar to several other measures of variability.

This paper considers the relationship between Δ and other measures of dispersion, including [1] the standard deviation, σ , [2] the average deviation about the mean, δ , [3] the range, R , and [4] quantile ranges such as the interquartile range, Q . Distribution-free inequalities are developed for the ratios of Δ to these other parameters. Also, the exact values of the ratios are listed for several population models and contrasted with the distribution-free results.

2. — DISTRIBUTION-FREE INEQUALITIES

For the ratio of the mean difference to the standard deviation of N variates

$$\frac{2}{\sqrt{N-1}} \leq \frac{\Delta}{\sigma} \leq \sqrt{\frac{4}{3} \frac{N+1}{N-1}} \quad [5]$$

and as $N \rightarrow \infty$

$$0 \leq \frac{\Delta}{\sigma} \leq \sqrt{\frac{4}{3}} \doteq 1.155 \quad [6]$$

both limits being attainable. The upper limit follows from the fact

$$\Delta = \frac{4}{N-1} \text{cov } i, X_{(i)} \leq \frac{4}{N-1} \sigma_i \sigma_X$$

and $\sigma_i^2 = (N^2 - 1)/12$. It is attained for any set of variates linearly related to the first N integers (rectangular distribution). The upper limit given is sharper than that sometimes quoted in the literature, namely $\sqrt{2}$. The lower limit can be developed by considering the maximum value of σ for a given Δ which occurs with, for example, a single value of $\mu - (N-1)\Delta/2$ and $N-1$ values of $\mu + \Delta/2$.

If $\mu > 0$, [5] and [6] also suggest limits for the ratio of Gini's coefficient of concentration, $\Delta/2\mu$, to the coefficient of variation, σ/μ . In particular, for $N \rightarrow \infty$, this ratio can not exceed $1/\sqrt{3} \doteq .577$.

For the ratio of the mean difference to the average (absolute) deviation about the mean

$$\frac{N}{N-1} \leq \frac{\Delta}{\delta} \leq 2 \quad [7]$$

so that δ can not exceed Δ , and Δ can not exceed 2δ . The limits follow from consideration of the minimum and maximum values of Δ given δ . The lower limit is attained for any and every binomial distribution. The upper limit is attained with single values of $\mu - N\delta/2$ and $\mu + N\delta/2$ and $N-2$ values of μ . (Chakravarti [1] gives comparable limits for δ/σ).

For the relationship between Δ and the various percentile ranges, one has, first of all

$$\begin{aligned} \Delta &\geq \frac{2(j-1)(N-j+1)}{N(N-1)} [X_{(j)} - X_{(i)}] & j > i \geq N-j+1 \\ &\geq \frac{2i(N-i)}{N(N-1)} [X_{(j)} - X_{(i)}] & N-i+1 \geq j > i \end{aligned} \quad [8]$$

This follows from considering the minimum value of Δ , given $X_{(j)} - X_{(i)}$ where $j > i$, which occurs with either i values of $X_{(i)}$ and $N-i$ values of $X_{(j)}$, or $j-1$ values of $X_{(i)}$ and $N-j+1$ values of $X_{(j)}$ depending on whether $i \leq N-j+1$ or $i \geq N-j+1$.

If $i=1$ and $j=N$ in [8], a lower limit to Δ/R is provided. An upper limit depends on whether N is even or odd and occurs when the distribution is equally, or nearly equally, divided at $X_{(1)}$ and $X_{(N)}$. Thus,

$$\begin{aligned} \frac{2}{N} \leq \frac{\Delta}{R} \leq \frac{1}{2} \frac{N}{N-1} & \quad (N \text{ even}) \\ \frac{2}{N} \leq \frac{\Delta}{R} \leq \frac{1}{2} \frac{N+1}{N} & \quad (N \text{ odd}) \end{aligned} \quad [9]$$

For other percentile ranges let ζ_p , the p -th quantile ($0 \leq p \leq 1$), equal $X_{(m)}$ where m is any integer satisfying

$Np \leq m \leq Np + 1$. Then from [8]

$$\Delta/(\zeta_p - \zeta_{1-p}) \geq 2 \frac{N}{N-1} p(1-p) \quad (p > 1/2) \quad [10]$$

which has $j = Np + 1$ and $i = N(1-p)$ and which may therefore be slightly improved if Np is not an integer. No upper limit to the ratio exists. For the interquartile range, Q , $p = .75$ in [10] and

$$\Delta/Q \geq \frac{3}{8} \frac{N}{N-1} \quad [11]$$

3. — PARAMETRIC RELATIONSHIPS

The inequalities given are distribution-free, but may be compared with exact results for various population models. For example, with a normal distribution with standard deviation, σ

$$\Delta = 2\sigma/\sqrt{\pi} \quad \delta = \sqrt{2/\pi} \sigma \quad Q = 1.349 \sigma \quad [12]$$

from which the values of Δ/σ , etc., can be determined. Similarly, for the rectangular distribution of range R

$$\Delta = R/3 \quad \sigma = R/\sqrt{12} \quad \delta = R/4 \quad Q = R/2 \quad [13]$$

and for the exponential distribution with mean μ ,

$$\Delta = \mu \quad \sigma = \mu \quad \delta = 2\mu/e \quad Q = (\log 3) \mu \quad [14]$$

Also if x has a Pareto distribution for $A \leq x \leq \infty$ and α is the Pareto coefficient

$$\begin{aligned} \Delta &= 2\alpha A/(\alpha-1)(2\alpha-1) & \delta &= 2A(\alpha-1)^{\alpha-2}/\alpha^{\alpha-1} \\ \sigma &= \sqrt{\alpha/(\alpha-2)} A/(\alpha-1) & Q &= [4^{1/\alpha} - (4/3)^{1/\alpha}] A \end{aligned} \quad [15]$$

but σ exists only if $\alpha > 2$ and Δ and δ exist only if $\alpha > 1$.

The following table summarizes numerical values of the ratios Δ/σ , Δ/δ , Δ/R , and Δ/Q , for these population models together with their distribution-free limits when $N \rightarrow \infty$.

	Δ/σ	Δ/δ	Δ/R	Δ/Q
Distribution-free (min)	0	1.000	0	.375
(max)	1.155	2.000	.500	—
Normal	1.128	1.414	0	.836
Rectangular	1.155	1.333	.333	.667
Exponential	1.000	1.359	0	.921
Pareto ($a = 1.5$)	0	1.299	0	2.293
($a = 3$)	.693	1.350	0	1.233

New York University

REFERENCE

(1) CHAKRAVARTI, M. C., «On the ratio of the mean deviation to the standard deviation,» *Calcutta Statistical Association Bulletin* (1947-48), 1, 187-190.

CARLO BENEDETTI

**A proposito dei rapporti tra differenza media
e scostamenti medio quadratico,
semplice medio e semplice medio dalla mediana**

1. La nota di J. Glasser ⁽¹⁾ che appare in questo stesso fascicolo di «Metron» e che, tra l'altro, tratta del campo di variazione dei rapporti Δ/σ , Δ/δ , Δ/Q , Δ/R (dove Δ = differenza media, σ = scarto quadratico medio, δ = scostamento semplice medio dalla media aritmetica, Q = differenza interquartile, R = campo di variazione), suggerisce di estendere a rapporti con altri indici la costruzione del rispettivo campo di variazione.

A dire il vero lo scopo principale di questa nostra breve nota avrebbe dovuto essere di pura applicazione statistica, ma poichè i dati di applicazione di cui disponiamo riguardano solo i rapporti già calcolati Δ/σ , Δ/δ , Δ/s , dove s è lo scostamento semplice medio dalla mediana, e non risultandomi ancora determinato il campo di variazione di Δ/s , lo abbiamo determinato in questa occasione principalmente per potere utilizzare tutti i dati statistici disponibili.

2. Data la successione di reali non decrescenti e non tutti uguali :

$$a_1 \leq a_2 \leq \dots \leq a_n \quad [1]$$

calcolando per essa Δ , σ , δ , s si ha :

$$\frac{2}{\sqrt{n-1}} \leq \frac{\Delta}{\sigma} \leq \sqrt{\frac{4}{3} \frac{n+1}{n-1}} \quad (1) \quad [2]$$

$$\frac{n}{n-1} \leq \frac{\Delta}{\delta} \leq 2 \quad (1) \quad [3]$$

⁽¹⁾ Relationship between the mean difference and other measures of variation in "Metron" vol. XXI n. 1-4, 1961.

e per quanto riguarda $\frac{\Delta}{s}$ si ha ancora :

$$\frac{n}{n-1} \leq \frac{\Delta}{s} \leq 2 \quad (\text{per } n \text{ pari}) \quad [4]$$

e :

$$\frac{n+1}{n-1} \leq \frac{\Delta}{s} \leq 2 \quad (\text{per } n \text{ dispari}) \quad [4']$$

dove si vede, per n pari, che si ha lo stesso intervallo di variazione di Δ/δ pur avendosi in generale $\frac{\Delta}{s} \geq \frac{\Delta}{\delta}$ per le note proprietà di minimo della mediana rispetto agli altri valori medi ; nella [4] si uguaglia il minimo nella successione :

$$a_1 = a_2 = \dots = a_{\frac{n}{2}} < a_{\frac{n}{2}+1} = a_{\frac{n}{2}+2} = \dots = a_n \quad [5]$$

mentre nella [4'] si ha il minimo nella successione :

$$a_1 = a_2 = \dots = a_{\frac{n-1}{2}} \leq a_{\frac{n+1}{2}} \leq a_{\frac{n+3}{2}} = \dots = a_n \quad [5']$$

Il massimo, sia per n pari che per n dispari, si ha nella successione :

$$a_1 \leq a_2 = \dots = a_{n-1} \leq a_n \quad [6]$$

3. Piuttosto banale è la prova della [4]. Tenendo presente la [1] si ha (a_m = mediana) :

$$\begin{aligned} n(n-1)\Delta &= \sum_{i \neq j} |a_i - a_j| = \sum_{i \neq j} |a_i - a_m + a_m - a_j| \leq \\ &\leq \sum_{i \neq j} \{ |a_i - a_m| + |a_m - a_j| \} = 2n(n-1)s \end{aligned}$$

da cui :

$$\Delta \leq 2s ; \quad \frac{\Delta}{s} \leq 2$$

cioè l'estremo destro della [4].

Inoltre si ha pure (tenendo presenti le note proprietà di minimo della mediana e ponendo $s_r = \frac{1}{n} \sum_{i=1}^n |a_i - a_r|$ con $1 \leq r \leq n$):

$$n(n-1)\Delta = \sum_{i \neq j} |a_i - a_j| = n \sum_{r=1}^n s_r \geq n^2 s$$

da cui:

$$\Delta \geq \frac{n}{n-1} s \quad ; \quad \frac{n}{n-1} \leq \frac{\Delta}{s}$$

cioè l'estremo sinistro della [4]. Si vede immediatamente che si tratta di minimo e di massimo in quanto tali estremi sono raggiunti da $\frac{\Delta}{s}$ rispettivamente nelle due successioni [5] e [6].

La prova della [4'], per quanto riguarda l'estremo destro è identica alla prova della [4]; per quanto riguarda invece l'estremo sinistro basta notare che con n dispari viene tolta l'ambiguità dei due termini centrali $\frac{a_n}{2}, \frac{a_{n+1}}{2}$ (del caso di n pari) a rappresentare la mediana ed allora possiamo scrivere, come si vede facilmente:

$$\sum_{i=1}^n |a_i - a_r| \geq n s + |a_r - a_m|^{(2)}$$

ossia:

$$n(n-1)\Delta = \sum_{r=1}^n \sum_{i=1}^n |a_i - a_r| \geq n^2 s + n s = n(n+1)s$$

da cui si ottiene l'estremo sinistro della [4'] che è raggiunto da Δ/s in [5']. Trattasi quindi, anche in questo caso, di minimo.

4. Dallo studio del Dott. Giovanni Dettori: *Contributo allo studio della variabilità dei prezzi*, Cagliari, 1912, si hanno i seguenti dati relativi ai rapporti calcolati per i prezzi di 50 merci in Italia dal 1879 al 1906 (50 serie di 28 termini ciascuna), per i prezzi di 20 merci in Francia dal 1857 al 1905 (20 serie

(2) Ricavabile agevolmente, ad esempio, dalla nota relazione ricorrente:

$$n s_{r+1} = n s_r + (2r-n)(a_{r+1} - a_r)$$

di 49 termini ciascuna) e per i salari della manodopera agricola tra i diversi dipartimenti del Regno Unito dal 1795 al 1893 :

CLASSI DEI VALORI DEI RAPPORTI	FREQUENZE DEI VALORI DEI RAPPORTI :								
	$\frac{\Delta}{\delta}$			$\frac{\Delta}{s}$			$\frac{\Delta}{\sigma}$		
	Italia	Francia	Inghilterra	Italia	Francia	Inghilterra	Italia	Francia	Inghilterra
0,60 - 0,70 . . .							1		
0,70 - 0,80 . . .							1		
0,80 - 0,90 . . .							—		
0,90 - 1,00 . . .							6		
1,00 - 1,10 . . .	1						8	2	2
1,10 - 1,20 . . .	4						25	17	6
1,20 - 1,30 . . .	11			1			9	—	
1,30 - 1,40 . . .	16	3		4	1			1	
1,40 - 1,50 . . .	16	6	4	9	11	3			
1,50 - 1,60 . . .			4	19	7	5			
1,60 - 1,70 . . .	2			11	1				
1,70 - 1,80 . . .				5					
1,80 - 1,90 . . .									
1,90 - 2,00 . . .				1					
	50	20	8	50	20	8	50	20	8
(valori teorici dei minimi e dei massimi dei rapporti)									
Minimi	1,0370	1,0208	(*)	1,037	1,0417	(*)	0,3849	0,2887	(*)
Massimi	2	2	2	2	2	2	1,1967	1,1785	(*)

(*) Non si conosce il numero dei termini delle serie e quindi non si possono determinare gli estremi del campo di variazione quando questi sono funzioni di n .

Dalla tabella precedente si nota che i rapporti Δ/σ calcolati dal Dettori sono, almeno in numero di dieci, errati e precisamente si tratta dei 9 rapporti per l'Italia che cadono nell'intervallo 1,20-1,30 e di 1 rapporto per la Francia che cade nell'intervallo 1,30-1,40 ; infatti essendo $\max \frac{\Delta}{\sigma} = 1,1967$ per l'Italia e $\max \frac{\Delta}{\sigma} = 1,1785$ per la Francia, è chiaro che nei 10 casi suddetti si sconfinava sicuramente dal campo di variazione.

A questo proposito bisogna tenere presente che nel 1912 (epoca alla quale risalgono i calcoli del Dettori) non essendoci i mezzi automatici di calcolo di oggi, gli errori di calcolo mate-

riale erano indubbiamente più probabili di oggi; forse i calcoli materiali dell'articolo in questione furono fatti addirittura a mano.

Comunque, accettando per buoni quei valori che cadono all'interno del proprio campo di variazione, si può dire che le distribuzioni di frequenza anche per un numero limitato di casi tendono a presentarsi unimodali, non molto asimmetriche e con una modesta dispersione.

Naturalmente, a rigore, le distribuzioni di frequenza riportate nella tabella non dovrebbero servire come preciso riferimento per lo studio delle vere e proprie distribuzioni teoriche di frequenza dei rapporti in questione, poichè nella tabella sono accomunati rapporti calcolati per merci diverse ed a questa diversità potrebbe imputarsi parte della dispersione dei rapporti intorno ai loro valori medi; comunque, dato che trattasi di rapporti tra indici di variabilità, cioè di grandezze nelle quali si possono presumere già eliminate le varie componenti differenziali delle caratteristiche delle varie merci (valori medi, ed in un certo modo, una parte della variabilità), queste distribuzioni che figurano nella tabella possono ugualmente ragguagliarci, a scanso di errori materiali di calcolo non individuabili, sul comportamento effettivo di tali rapporti nel caso concreto di un numero finito di essi e qualora vengano calcolati per serie statistiche con numero finito di termini.

SUMMARY

The Author takes the hint from the note by J. Glasser that appears in this same number of «Metron» and he carries out the range of the ratio Δ/s (Δ = mean difference and s = mean simple deviation about the median of a succession of n real terms not all equal), that is:

$$\frac{n}{n-1} \leq \frac{\Delta}{s} \leq 2; \quad (n \text{ even})$$

$$\frac{n+1}{n-1} \leq \frac{\Delta}{s} \leq 2; \quad (n \text{ odd})$$

Besides, many statistical applications about the distribution of the ratios Δ/σ and Δ/δ (σ = standard deviation, δ = mean simple deviation about the mean), already appeared in Glasser's note, and of the ratio Δ/s , have been made.

Le tabelle di co- e contrograduazione e la programmazione lineare ⁽¹⁾

AMATO HERZEL

SOMMARIO: 1. Le tabelle di cograduazione e di contrograduazione (tabelle del Salvemini): costruzione e scopi originari. Le tabelle come algoritmo per la soluzione del « problema dei trasporti » in determinati casi. 2. Condizioni della loro applicabilità. 3. Casi particolari di applicabilità. Esempificazione. 4. Estensione a un problema più generale. 5. Deduzioni riguardanti la funzione minima e massima della classe di Fréchet.

1. Date due distribuzioni, egualmente numerose, di caratteri quantitativi, è noto che, se si associa ad ogni modalità del primo carattere una modalità del secondo, la somma dei quadrati delle differenze fra le modalità associate è massima quando le modalità sono contrograduate e minima quando sono cograduate. Su questa proprietà si fondano, tra l'altro, gli indici di omofilia, di dissomiglianza e di connessione introdotti nella metodologia statistica da C. Gini.

Per rendere praticamente possibile il calcolo di detti indici, il Salvemini ha ideato²⁾ le cosiddette tabelle di cograduazione e contrograduazione che consentono di ricavare, con un procedimento semplicissimo, rispettivamente la distribuzione doppia minimante e quella massimante la funzione suddetta.

Sia X il I° carattere: denotiamo le sue modalità disposte in ordine crescente con x_1, x_2, \dots, x_h e le frequenze corrispondenti con N_1, N_2, \dots, N_h . Siano Y il secondo carattere, y_1, y_2, \dots, y_k , le mo-

(1) Lavoro presentato al Seminario di Statistica metodologica dello Istituto di Statistica della Facoltà di Scienze Statistiche, Demografiche ed attuariali dell'Università di Roma e discusso nella riunione del 27 aprile 1957.

(2) T. SALVEMINI, *Sugli indici di omofilia* in « Atti della 1 Riunione della Società Italiana di Statistica », Roma, 1939.

dalità, parimenti in ordine crescente e N'_1, N'_2, \dots, N'_k le rispettive frequenze. Per ipotesi si ha

$$\sum_{i=1}^h N_i = \sum_{j=1}^h N_j$$

Il procedimento di Salvemini consiste nel costruire con questi dati determinate tabelle a doppia entrata.

In pratica si opera come segue: si scrivono nella prima colonna le modalità di un carattere e nell'ultima colonna le frequenze corrispondenti ed analogamente si scrivono nella prima riga le modalità del secondo carattere e nell'ultima riga le frequenze corrispondenti, come dal seguente schema:

$X \backslash Y$	y_1	y_2	y_k	
x_1					N_1
x_2					N_2
...					...
x_h					N_h
	N'_1	N'_2	N'_k	

Per le tabelle di *cograduazione* (che permettono di determinare la distribuzione doppia minimante) si considera dapprima la casella corrispondente alle modalità x_1, y_1 , in cui si scrive la più piccola fra le due frequenze N_1 e N'_1 , se $N_1 > N'_1$ si completa la prima colonna con altrettanti zeri e si scrive nella seconda casella della prima riga la più piccola fra le quantità $N_1 - N'_1$ e N'_2

e così via, alternativamente per righe e colonne. Se $N_1 < N'_1$, si completa invece la prima riga con altrettanti zeri, si scrive nella seconda casella della prima colonna la più piccola delle quantità $N'_1 - N_1$ e N_2 e così via.

Per le tabelle di *contrograduazione* (che forniscono la distribuzione doppia massimante) si procede in maniera del tutto analoga, solo che si comincia dalla casella corrispondente alle modalità x_1, y_k (o x_h, y_1).

Per rendere più evidente quanto esposto, facciamo seguire un semplice esempio numerico, con $h = 3, k = 4$:

Tabella di cograduazione.

$X \backslash Y$	y_1	y_2	y_3	y_4	
x_1	5	3	2	0	10
x_2	0	0	2	6	8
x_3	0	0	0	2	2
	5	3	4	8	20

Tabella di contrograduazione

$X \backslash Y$	y_1	y_2	y_3	y_4	
x_1	0	0	2	8	10
x_2	3	3	2	0	8
x_3	2	0	0	0	2
	5	3	4	8	20

Il significato di queste tabelle è chiaro: fra tutte le distribuzioni doppie aventi le distribuzioni marginali uguali a quelle assegnate, quella che minimizza la somma dei quadrati delle differenze fra le modalità associate presenta la coppia x_1, y_1 con frequenza 5, la coppia x_1, y_2 con la frequenza 3 e così via; quella che massimizza detta somma presenta la coppia x_1, y_4 con frequenza 8, la coppia x_1, y_3 con frequenza 2 e così via.

L'importanza delle tabelle del Salvemini per lo studio delle relazioni fra variabili statistiche, ha trovato vasti quanto meriti riconosciuti. Non risulta però che siano stati finora fatti dei tentativi per indagarne il significato matematico più generale e astratto (svincolato, cioè, dal problema specifico in funzione del quale sono state ideate). Questa generalizzazione costituisce appunto lo scopo della presente nota.

Consideriamo anzitutto le N_i ($i = 1, 2, \dots, h$) e le N'_j ($j = 1, 2, \dots, k$) quantità positive qualunque, cioè non più necessariamente frequenze. Rimane naturalmente ferma l'eguaglianza:

$$\sum_{i=1}^h N_i = \sum_{j=1}^k N'_j$$

Denotiamo con λ_{ij} la variabile la cui determinazione viene iscritta nella casella corrispondente alla coppia x_i, y_j nella tabella di cograduazione o di contrograduazione. Le λ sono quantità non negative che devono soddisfare alle seguenti condizioni:

$$\sum_{j=1}^k \lambda_{ij} = N_i \quad (i = 1, 2, \dots, h) \quad [1]$$

$$\sum_{i=1}^h \lambda_{ij} = N'_j \quad (j = 1, 2, \dots, k) \quad [2]$$

Possiamo inoltre simbolizzare nel seguente modo la somma dei quadrati delle differenze fra le modalità corrispondenti:

$$\varphi = \sum_{i,j} \alpha_{ij} \lambda_{ij}$$

dove $\alpha_{ij} = (x_i - y_j)^2$, ($x_1 < x_2 < x_3, \dots; y_1 < y_2 < y_3 \dots$) (3)

Nella tabella di *cograduazione* le λ vengono determinate in modo da rendere minima la φ , nelle tabelle di *contrograduazione* in modo da rendere massima tale funzione.

Si vede immediatamente che le tabelle di cograduazione, considerate sotto questo aspetto, altro non sono che un algoritmo per risolvere determinati problemi di programmazione lineare, anzi, per essere più precisi, esse, sotto certe condizioni che specificheremo in seguito, costituiscono un metodo per risolvere il cosiddetto « problema dei trasporti » o di Hitchcock - Koopmans (1). Questo « problema » trova applicazioni svariatissime che vanno ben al di là dell'impostazione originaria, alla quale deve la sua de-

(1) La letteratura sulla programmazione lineare e anche sul « problema dei trasporti » si è così estesa in questi ultimi anni che non ci è possibile fornire una bibliografia anche approssimativamente esauriente. Ci limitiamo perciò a segnalare al lettore che volesse approfondire l'argomento, la pubblicazione *La programmazione lineare nell'industria* (Torino, 1954), che contiene anche una ricca bibliografia e una recente nota di M. SALVADORI *La programmazione lineare* («La civiltà delle macchine» Gennaio - Febbraio 1957).

nominazione. Una merce si trova accumulata in certe località e deve essere trasportata in altre; sono dati i quantitativi delle merci accumulate in ogni località di partenza, quelli che devono giungere in ogni località di arrivo e i costi unitari per i trasporti da ogni località di partenza ad ogni località di arrivo. Si tratta di effettuare i trasporti in modo tale da rendere minimo il costo complessivo.

2. Il metodo di soluzione, dovuto al Dantzig, si può così riassumere: si dispongono i termini noti (che corrispondono ai nostri N_i e N'_j) ordinatamente ai margini di una tabella a doppia entrata e si costruisce quindi la « soluzione-base » che *formalmente* (in quanto non si tiene conto dei pesi α_{ij}) (1) è uguale alla tabella di cograduazione. Le variabili che nella soluzione — base assumono un valore diverso da 0 (e quindi positivo) vengono denominate variabili-base. Il loro numero è normalmente di $h + k - 1$ (non può essere maggiore) e noi supporremo in seguito che sia tale. I successivi svolgimenti rimangono tuttavia validi anche se il numero delle variabili che assumono un valore positivo è minore.

In base ad un noto principio della programmazione lineare, per cui si giunge alla soluzione ottima attraverso successivi miglioramenti, occorre ora vedere se è possibile ottenere un risultato migliore scrivendo in una delle caselle rimaste vuote (contenenti, cioè, uno zero) una cifra significativa, ossia facendo assumere ad una delle λ , che nella soluzione-base hanno assunto la determinazione 0, una determinazione positiva. Poichè le somme per righe e per colonne sono date, se si pone un'unità in una casella, occorre diminuire di una unità il valore assunto da una variabile-base posta sulla stessa riga (colonna), aumentare di un'unità quella di una variabile-base posta sulla stessa colonna (riga) di quest'ultima e così via, fino a diminuire di una unità il valore di una variabile-base posta sulla stessa colonna (riga) di quella di partenza. Per misurare l'effetto che questo cambiamento produce sulla funzione: $\sum \alpha_{ij} \lambda_{ij}$ che deve essere resa minima, basta fare la somma algebrica dei pesi α_{ij} corrispondenti alle caselle che dovrebbero subire variazioni, prenden-

(1) Qui, e in seguito, le α_{ij} sono costanti generiche, che non soddisfano necessariamente le relazioni (3).

do con segno positivo quelli a cui corrisponderebbero aumenti e con segno negativo quelli a cui corrisponderebbero diminuzioni. Se la somma è positiva, il cambiamento non è conveniente, lo è invece se la somma è negativa.

Si ripete questa operazione per tutte le caselle vuote (nel caso della nostra tabella di cograduazione si avrebbe quindi successivamente: $\alpha_{14} - \alpha_{13} + \alpha_{23} - \alpha_{24}$, $\alpha_{21} - \alpha_{23} + \alpha_{13} - \alpha_{11}$, $\alpha_{22} - \alpha_{23} + \alpha_{13} - \alpha_{11}$, $\alpha_{31} - \alpha_{34} + \alpha_{24} - \alpha_{23} + \alpha_{13} - \alpha_{11}$, $\alpha_{32} - \alpha_{34} + \alpha_{24} - \alpha_{23} + \alpha_{14} - \alpha_{12}$, $\alpha_{33} - \alpha_{34} + \alpha_{24} - \alpha_{23}$).

Se tutte queste somme algebriche dei pesi, che indicheremo con S_{ij} (se λ_{ij} è la variabile da cui si parte) sono non-negative, la soluzione iniziale è anche quella ottima; se una od alcune sono negative, bisogna far assumere alla variabile che apporterebbe il massimo miglioramento (alla quale, cioè, corrisponde la minima S_{ij} negativa) il massimo valore possibile, in modo da annullare una variabile-base. Si ripete poi il procedimento fino a quando ogni nuova modifica comporterebbe un peggioramento.

È evidente che questa procedura è grandemente semplificata quando la soluzione base è quella ottima. Di qui l'importanza pratica della questione: a quali condizioni devono soddisfare i pesi α_{ij} affinché la tabella di cograduazione (o soluzione-base) fornisca la soluzione ottima, quali che siano le quantità N_i e N'_j ?

Perchè ciò avvenga, è necessario e sufficiente che siano verificate le seguenti relazioni:

$$\alpha_{ij} + \alpha_{il} \leq \alpha_{il} + \alpha_{lj} \quad [4]$$

per:

$$(i = 1, 2, \dots, h-1; j = 1, 2, \dots, k-1; \\ t = i+1, i+2, \dots, h; l = j+1, j+2, \dots, k)$$

La condizione è *necessaria*; poniamo, infatti, che si abbia per esempio:

$\alpha_{mn} + \alpha_{pq} > \alpha_{mq} + \alpha_{pn}$ ($p > m$, $q > n$). Per certi valori delle N e N' si avrà nella soluzione base: $\lambda_{m,n} > 0$, $\lambda_{n,p} < 0$, $\lambda_{p,q} > 0$; $\lambda_{m,q} = 0$, e $S_{m,q} = \alpha_{m,q} - \alpha_{m,n} + \alpha_{p,n} - \alpha_{p,q} < 0$, e quindi la soluzione-base non sarà quella ottima.

La condizione è anche *sufficiente*: basta osservare che, data la struttura della tabella di cograduazione, quando in una casella si trova uno zero, se la casella più vicina della stessa riga

recante una cifra significativa è collocata a destra della casella vuota, la casella più vicina della stessa colonna recante una cifra significativa si trova in alto; viceversa se la casella più vicina recante una cifra significativa si trova a sinistra della casella vuota, quella più vicina sulla stessa colonna contenente una cifra significativa si trova in basso. Poniamoci in questo ultimo caso. Sia $\lambda_{m,q} = 0$, e il percorso da compiere per verificare se la soluzione possa essere migliorata, ponendo $\lambda_{mq} > 0$, tocchi successivamente le caselle corrispondenti alle variabili $\lambda_{m,q} \lambda_{m,n} \lambda_{m_1,n} \dots \lambda_{m_r,n_r} \lambda_{p,n_r} \lambda_{p,q}$ ($m < m_1 < m_2, \dots, < m_r < p; n < n_1 < n_2 \dots < n_r < q$).

Si dovrà quindi calcolare l'espressione:

$$S_{m,q} = \alpha_{m,q} - \alpha_{m,n} + \alpha_{m_1,n} - \alpha_{m_1,n_1} + \dots - \alpha_{m_r,n_r} + \alpha_{p,n_r} - \alpha_{p,q}$$

Questa espressione si può anche scrivere:

$$S_{mq} = (\alpha_{mq} - \alpha_{mn} + \alpha_{m_1,n} - \alpha_{m_1,q}) + (\alpha_{m_1,q} - \alpha_{m_1,n_1} + \alpha_{m_2,n_1} - \alpha_{m_2,q}) + \dots + (\alpha_{m_r,q} - \alpha_{m_r,n_r} + \alpha_{p,n_r} - \alpha_{p,q}).$$

Ma tutte le espressioni in parentesi di questa somma sono o positive o nulle se sono verificate le relazioni [4]. In modo perfettamente analogo si dimostra che ciò vale anche per le somme relative alle caselle nulle aventi le caselle con cifre significative in alto e a destra. Ciò dimostra che la condizione enunciata è anche sufficiente, affinché la soluzione-base non sia suscettibile di miglioramenti.

Si dimostra ugualmente che, nel caso che valga tale disuguaglianza, la tabella di contrograduazione fornisce il valore massimo della funzione: $\varphi = \sum \alpha_{ij} \lambda_{ij}$, mentre viceversa, qualora si abbia sempre:

$$\alpha_{ij} + \alpha_{il} \geq \alpha_{il} + \alpha_{ij} \quad (i < t \leq h, j < l \leq k) \quad [5]$$

la tabella di cograduazione rende massima la funzione suddetta e quella di contrograduazione la rende minima.

3. Le condizioni enunciate (che in pratica si equivalgono perchè supposto verificate per es. le disuguaglianze [4] basta invertire l'ordine delle righe o delle colonne per passare alle disu-

guaglianze opposte), sono meno restrittive di quanto possa apparire a prima vista. Per esempio, esse possono essere sempre soddisfatte, eventualmente attraverso un opportuno spostamento delle righe e delle colonne, quando si abbia $\alpha_{ij} = |x_i - y_j|^r$ con x_i e y_j quantità reali arbitrarie e r intero e positivo, o $\alpha_{ij} = |x_i + y_j|^r$, o ancora $\alpha_{ij} = x_i y_j$.

Più precisamente, se si ha : $x_i < x_t$, $y_j < y_l$ ($i < t$, $j < l$) e se i pesi sono della forma :

$$\alpha_{ij} = |x_i - y_j|^r$$

con r intero e positivo, si verificano costantemente le disuguaglianze [4] e quindi il minimo è fornito dalla tabella di cograduazione e il massimo da quella di contrograduazione ; se sono invece della forma :

$$\alpha_{ij} = |x_i + y_j|^r$$

$$\text{oppure : } \alpha_{ij} = x_i y_j$$

sono verificate costantemente le disuguaglianze [5] e il minimo della funzione $\varphi = \sum \alpha_{ij} \lambda_{ij}$ è dato dal sistema delle λ fornito dalla tabella di contrograduazione mentre il massimo corrisponde alla tabella di cograduazione..

Ancora : le tabelle di cograduazione o contrograduazione forniscono la soluzione ottima quando i pesi possono essere disposti in modo da variare in progressione aritmetica nelle singole righe o nelle singole colonne. Poniamo, per esempio, che siano in progressione aritmetica nel senso delle righe. Potremo allora scrivere :

$$\alpha_{ij} = x_i + (j - 1) d_i$$

dove x_i rappresenta il primo termine e d_i la ragione della progressione aritmetica dei pesi posti nella riga i esima. Se spostiamo allora le righe, in modo da disporle in ordine non crescente delle d , avremo :

$$\alpha_{i'j} = x_{i'} + (j - 1) d_{i'}$$

e avremo inoltre :

$$d_{i'} \geq d_{i''}, \quad (\text{per } t' > i').$$

Ne consegue :

$$\alpha_{i'l} + \alpha_{i''l} = x_{i'} + (j-1) d_{i'} + x_{i''} + (l-1) d_{i''}$$

$$\alpha_{i'l} + \alpha_{i''j} = x_{i'}' + (l-1) d_{i'} + x_{i''} + (j-1) d_{i''}$$

e quindi :

$$\alpha_{i'l} + \alpha_{i''j} - (\alpha_{i'l} + \alpha_{i''l}) = (d_{i'} - d_{i''}) (l-j) \geq 0$$

cioè :

$$\alpha_{i'l} + \alpha_{i''j} \leq \alpha_{i'l} + \alpha_{i''l} \quad (\text{per } t' > i', l > j)$$

e il minimo corrisponde pertanto alla tabella di cograduazione, mentre il massimo è fornito dalla tabella di contrograduazione. Ne discende, come conseguenza immediata che la notevole semplificazione dei calcoli, insita nell'impiego delle tabelle di Salvemini, potrà sempre trovare applicazione quando le righe (colonne) siano soltanto due, qualunque sia il numero delle colonne (righe); per trovare il sistema minimante basterà disporre le colonne (righe) in modo che le differenze fra i pesi della prima e della seconda riga (colonna) siano in ordine non decrescente (per le tabelle di graduazione), o non crescente (per le tabelle di contrograduazione).

Per illustrare questo procedimento, prendiamo un esempio semplicissimo, tratto dalla citata nota di M. Salvadori.

La tabella dei pesi α_{ij} è la seguente :

	1	2	3
1	2	1	3
2	1	4	2

Si tratta di rendere minima la funzione . $\sum \alpha_{ij} \lambda_{ij}$, subordinatamente alle condizioni .

$$\begin{aligned}\lambda_{11} + \lambda_{12} + \lambda_{13} &= 50 \\ \lambda_{21} + \lambda_{22} + \lambda_{23} &= 80 \\ \lambda_{11} + \lambda_{31} &= 60 \\ \lambda_{12} + \lambda_{22} &= 30 \\ \lambda_{13} + \lambda_{23} &= 40 \\ \lambda_{ij} &\geq 0 \quad (i = 1, 2 ; j = 1, 2, 3)\end{aligned}$$

Si verifica immediatamente che possiamo disporre le colonne indifferentemente nell'ordine 2, 1, 3 o 2, 3, 1 ottenendo così le seguenti tabelle di cograduazione :

	2	1	3			2	3	1	
1	30	20		50	1	30	20		50
2		40	40	80	2		20	60	80
	30	60	40			30	40	60	

che ci forniscono due soluzioni ottime, senza bisogno di ulteriori controlli, che sarebbero invece stati necessari se avessimo applicato il metodo generale per la soluzione del « problema dei trasporti ».

Anche quando, nel problema dei trasporti in senso stretto, le località di partenza da un lato e quelle di arrivo dall'altro sono approssimativamente allineate e i costi possono essere considerati proporzionali alle distanze, il metodo delle tabelle di cograduazione può portare ad una rapida soluzione. Ecco un esempio.

Un'impresa industriale abbia stabilimenti a Genova, Napoli, Roma e Torino e desideri portare i suoi prodotti nei seguenti centri : l'Aquila, Firenze, Milano, Perugia, Potenza. Supposto che i costi unitari dei trasporti siano proporzionali alle distanze in linea d'aria, si avra una tabella dei costi del seguente tipo :

	Genova	Napoli	Roma	Torino
L'Aquila	426	177	90	546
Firenze.	198	402	231	318
Milano	117	651	477	123
Perugia.	312	291	135	432
Potenza	690	129	306	810

Uno sguardo alla carta geografica permette di constatare che i centri di produzione sono approssimativamente allineati, e così i centri di smercio. Se si ordinano entrambi da Nord a Sud, la tabella dei costi assume il seguente aspetto :

	Torino	Genova	Roma	Napoli
Milano	123	117	477	655
Firenze	318	198	231	402
Perugia	432	312	135	291
L'Aquila	546	426	90	177
Potenza	810	690	306	129

Com'è facile verificare, sono qui pienamente soddisfatte le condizioni per l'applicabilità del metodo della tabella di cograduazione.

Supposto che i singoli centri di produzione dispongano di un numero di unità del prodotto per la spedizione come in appresso indicato :

	unità da spedire
Genova	1.000
Napoli	800
Roma	1.500
Torino	2.000
Totale	5.300

e supposto che i singoli centri di smercio possano assorbire le quantità qui specificate :

	unità da ricevere
L'Aquila	100
Firenze	1.200
Milano	3.500
Perugia	200
Potenza	300
Totale	5.300

il problema si risolve immediatamente mediante la tabella di cograduazione come segue :

	Torino	Genova	Roma	Napoli	Totali da ricevere
Milano	2.000	1.000	500		3.500
Firenze.			1.000	200	1.200
Perugia				200	200
L'Aquila				100	100
Potenza				300	300
Totali da spedire	2.000	1.000	1.500	800	

Milano riceverà quindi 2.000 unità del prodotto da Torino, 1.000 da Genova e 500 da Roma, Firenze riceverà 1.000 unità da Roma e 200 da Napoli, che coprirà anche tutto il fabbisogno di Perugia, L'Aquila e Potenza.

4. Un altro problema, in sostanza analogo, ma alquanto più generale, nel quale possono trovare applicazione le tabelle del Salvemini, è il seguente :

Sia da rendere minima la funzione :

$$\sum \alpha_{ij} \lambda_{i,j},$$

coi seguenti vincoli :

$$\begin{aligned} \sum_j \beta_{i,j} \lambda_{i,j} &= N_i \\ \sum_i \beta_{i,j} \lambda_{i,j} &= N'_j \end{aligned} \quad (\sum N_i = \sum N'_j),$$

dove le $\beta_{i,j}$ le N_j e le N' , sono costanti positive arbitrarie e le λ non possono assumere valori negativi. Poniamo :

$$\gamma_{ij} = \frac{\alpha_{i,j}}{\beta_{i,j}}$$

Se tutte le γ soddisfano alla condizione :

$$\gamma_{i,j} + \gamma_{t,l} \leq \gamma_{i,l} + \gamma_{t,j} \quad (i < t \leq h, j < l \leq k)$$

possiamo determinare mediante la tabella di cograduazione i valori delle variabili $\gamma_{i,j}$ che rendono minima la funzione:

$$\sum \gamma_{i,j} \mu_{i,j}$$

coi seguenti vincoli:

$$\mu_{i,j} \geq 0$$

$$\sum_j \mu_{i,j} = N_i$$

$$\sum_i \mu_{i,j} = N'_j$$

La soluzione ottima del primo sistema è allora data, come si vede facilmente, dalle determinazioni:

$$\lambda'_{i,j} = \frac{\mu'_{i,j}}{\beta_{i,j}}$$

dove le $\mu'_{i,j}$ sono le determinazioni delle variabili $\mu_{i,j}$ fornite dalla tabella di cograduazione.

5. I risultati a cui siamo giunti sono suscettibili di essere trasposti anche nella teoria statistica vera e propria. A questo scopo torniamo al punto di partenza della nostra indagine. Come abbiamo visto, le tabelle di Salvemini determinano nel discontinuo certe distribuzioni doppie, essendo assegnate le distribuzioni semplici delle due variabili prese separatamente. Questa impostazione è stata generalizzata, con risultati interessanti, da M. Fréchet (1) in modo da estendersi anche al continuo.

Date le due funzioni di ripartizione $\psi_1(x)$ e $\psi_2(y)$ delle due variabili statistiche X e Y , è chiamato classe di Fréchet (2)

(1) M. FRÉCHET, *Sur les tableaux de corrélation dont les marges sont données*, « Annales de l'Université de Lyon », III fasc. 4: « Sciences », 1951. Dello stesso Autore *Les tableaux de corrélation et les programmes linéaires*, in « Revue de l'Institut International de Statistique », vol. 25, n. 1-3, l'Aja 1957. Tale articolo è stato pubblicato quando il presente lavoro era già stato completato (e presentato al Seminario di Statistica metodologica dell'Istituto di Statistica dell'Università di Roma).

(2) Cfr. G. DALL'AGLIO, *Gli estremi dei momenti delle funzioni di ripartizione doppie*, « Annali della Scuola Normale Superiore » 1956; G. LANDENNA, *La dissomiglianza tra variabili casuali*, « Statistica », 1956; G. POMPILJ, *Teoria affine delle variabili casuali*, « L'Industria », 1956.

l'insieme delle funzioni di ripartizione doppia, associate alle due variabili, che soddisfano alle seguenti condizioni :

$$\lim_{y \rightarrow \infty} F(x, y) = \psi_1(x)$$

$$\lim_{x \rightarrow \infty} F(x, y) = \psi_2(y)$$

Nella classe di Fréchet assumono particolare importanza due funzioni :

$$F^0(x, y) = \frac{1}{2} [\psi_1(x) + \psi_2(y) - 1 + |\psi_1(x) + \psi_2(y) - 1|]$$

$$F^1(x, y) = \frac{1}{2} [\psi_1(x) + \psi_2(y) - |\psi_1(x) - \psi_2(y)|]$$

La $F^0(x, y)$ e la $F^1(x, y)$ assumono in corrispondenza ad ogni coppia di valori reali (x, y) rispettivamente il minimo e il massimo valore compatibile con l'appartenenza alla classe e vengono perciò chiamate funzione minima e funzione massima. La funzione minima coincide nel discontinuo con la distribuzione doppia determinata dalla tabella di contrograduazione, la massima con quella determinata dalla tabella di cograduazione.

Ciò premesso, possiamo porci il seguente problema, sostanzialmente analogo a quello risolto nella prima parte del presente lavoro, limitatamente al caso discontinuo :

Siano date una funzione reale in due variabili, $g(x, y)$, del tutto arbitraria e le funzioni di ripartizione semplice $\psi_1(x)$ e $\psi_2(y)$. Nella classe di Fréchet, definita da queste ultime funzioni, sia $F'(x, y)$ la funzione di ripartizione doppia che rende minimo (massimo) l'integrale di Stieltjes :

$$J = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) dF(x, y)$$

In generale, $F'(x, y)$ varierà al variare della funzione $g(x, y)$. Possiamo quindi chiederci: a quali condizioni deve soddisfare la $g(x, y)$, affinchè la funzione $F'(x, y)$, ammesso che esista, coincida con la funzione massima (minima) della classe di Fréchet, comunque si scelgano le funzioni di ripartizione $\psi_1(x)$ e $\psi_2(y)$?

Da quanto abbiamo dimostrato in precedenza, risulta che condizione *necessaria* perchè ciò avvenga è che in tutto il piano (x, y) sia verificata la seguente disuguaglianza.

$$g(x, y) + g(x + \Delta x, y + \Delta y) \leq g(x, y + \Delta y) + g(x + \Delta x, y) \\ [g(x, y) + g(x + \Delta x, y + \Delta y) \geq g(x, y + \Delta y) + g(x + \Delta x, y)]$$

in altri termini, tutte le variazioni doppie della $g(x, y)$ devono risultare negative (positive) o nulle. In tal caso è verificata una condizione necessaria perchè la funzione massimante sia rappresentata dalla funzione minima (massima) della classe di Fréchet.

Queste condizioni sono anche *sufficienti*? Per analogia con quanto si verifica nel discontinuo, si è indotti a rispondere affermativamente; tuttavia, a quanto ci risulta, manca ancora una dimostrazione rigorosa.

RIASSUNTO

Si dimostra che le tabelle di co- e contrograduazione, impiegate in statistica per il calcolo degli indici di dissomiglianza e di connessione, rappresentano un algoritmo per risolvere un tipo di problemi di programmazione lineare (denominato «problema dei trasporti» o «problema di Hitchcock»), a condizione che i coefficienti della funzione lineare che deve essere resa massima o minima soddisfino un determinato sistema di disuguaglianze. Questo risultato, da un lato, può essere utilizzato per una più facile e rapida soluzione di alcuni problemi di programmazione lineare e dall'altro lato serve a mettere in luce alcune caratteristiche delle tabelle di co- e contrograduazione.

GIUSEPPE LETI

Le distribuzioni degli estremi e del campo di variazione dei campioni di una popolazione discreta e finita (*).

1. — Negli ultimi anni l'estremo inferiore, l'estremo superiore e il campo di variazione sono stati rivalutati, nella statistica, dalla teoria dei campioni, in quanto forniscono stime della variabilità che, pur essendo alquanto grossolane, sono generalmente poco costose e rapide e quindi assai apprezzate da chi deve servirsi tempestivamente delle informazioni fornite dai campioni.

È per questa ragione che numerosi studi sono stati dedicati a tali caratteristiche di posizione ; ma è da notare che, pur essendo state considerate popolazioni distribuite in modo qualunque, la unica operazione campionaria presa in esame è stata l'estrazione bernoulliana. Fa eccezione il lavoro di T. Salvemini⁽¹⁾ che ha determinato la legge di distribuzione del campo di variazione nell'universo campionario generato dall'operazione di estrazione in blocco, operante su una popolazione equidistribuita.

Poichè oltre alla operazione di estrazione bernoulliana ve ne sono altre da prendere in considerazione, in questo lavoro sono state determinate le leggi di distribuzione degli estremi e del

(*) Parte di questo lavoro, e precisamente quella che si riferisce allo universo in blocco, è stata comunicata al Seminario di Statistica Metodologica presso la Facoltà di Scienze Statistiche Demografiche ed Attuariali dell'Università degli Studi di Roma, nella riunione del 29 marzo 1958, mentre la parte che riguarda la trattazione generale è stata comunicata il 14 settembre 1959 al VI Congresso Nazionale dell'Unione Matematica Italiana a Napoli.

(¹) SALVEMINI T.: *Sui campioni di una massa discreta equidistribuita: il campo di variazione*. « Statistica » anno XVIII n. 1 ; gennaio-marzo 1958.

campo di variazione in universi generati da generali operazioni campionarie, operanti su popolazioni discrete, finite e distribuite in modo qualunque.

Le formule ottenute sono state particolarmente studiate per gli universi generati dalle operazioni di estrazione con ripetizione e senza ripetizione e, mentre le formule relative all'universo bernoulliano erano già conosciute, sebbene ottenute con difficoltà e non facilmente generalizzabili, quelle dell'universo in blocco non erano fino ad ora note.

LA POPOLAZIONE ⁽²⁾

2. — La popolazione **P** è l'insieme totale delle N unità che formano l'oggetto della rivelazione.

Le unità della popolazione siano di natura qualunque e siano raggruppate in $m + 1$ classi rispetto ad un carattere X , anch'esso di natura qualsiasi. Siano

$$x_0, x_1, \dots, x_{m-1}, x_m \quad [1]$$

i simboli con cui rappresentiamo le classi e che chiamiamo modalità del carattere X , e

$$n_0, n_1, \dots, n_{m-1}, n_m \quad [2]$$

il numero delle unità di **P** comprese rispettivamente nelle classi

$$x_0, x_1, \dots, x_{m-1}, x_m$$

con

$$n_0 + n_1 + \dots + n_{m-1} + n_m = N.$$

⁽²⁾ La nomenclatura seguita è quella adottata dal CASTELLANO in: CASTELLANO V.: *Cours de sondage: théorie et méthodes*. Organisation des Nations Unies pour l'Alimentation et l'Agriculture. Roma, 1955; ovvero in CASTELLANO V.: *Sull'universo dei campioni come fondamento della teoria dei campioni*. «Bull. de l'Institut International de Statistique», t. XXXVII, 1960.

Finchè sui simboli $[1]$ non vengono introdotte ipotesi di alcun genere, essi costituiscono un insieme di elementi sul quale non è possibile fare considerazioni di sorta.

In questo lavoro ci limitiamo ad introdurre il minimo indispensabile di ipotesi sui simboli $[1]$: così, poichè per parlare di estremi è necessario supporre che fra le modalità della popolazione si possa definire un ordinamento, noi supponiamo che sia

$$x_0 < x_1 < x_2 < \dots < x_{m-1} < x_m. \quad [3]$$

Per il campo di variazione, oltre all'ipotesi fatta, è necessario supporre che fra due modalità, la prima maggiore della seconda, si possa definire l'operazione di sottrazione, che goda delle proprietà formali, che, nell'aritmetica, caratterizzano questa operazione.

Per definire infine la media aritmetica, introduciamo l'ipotesi che sia possibile l'operazione di addizione, che goda anche essa delle proprietà formali di cui gode, nell'aritmetica, l'addizione.

L'UNIVERSO DEI CAMPIONI.

3. — Un operazione è detta « campionaria di ordine v » (v intero non negativo) se, applicata alla popolazione P , dà per risultato un gruppo di v unità di P , non necessariamente distinte, e se i risultati che fornisce sono tutti egualmente possibili.

Chiamiamo « universo dei campioni di ordine v », e lo indichiamo con U_v , l'insieme di tutti i campioni di v elementi che sono possibili in base all'operazione campionaria, di ordine v , operante sulla popolazione P .

Per la trattazione che seguirà supponiamo che l'universo U , sia tale che si conosca la funzione

$$u = f(t_0, t_1, t_2, \dots, t_m, v)$$

nella variabile v e nelle $m + 1$ variabili t_h , la quale, comunque si fissino due indici i e k ($0 \leq i \leq k$; $i, k = 0, 1, \dots, m$), fornisca il numero di tutti i campioni di v unità formati con le sole modalità $x_i, x_{i+1}, \dots, x_{k-1}, x_k$, allorchè è $t_h = 0$, per $h < i$ e per $h > k$, e $t_h = n_h$, per ogni h non inferiore ad i e non superiore a k .

Tale funzione, che chiamiamo *funzione di struttura* dell'universo campionario, la indichiamo brevemente con

$$f_{ik}(v) = f(n_i, n_{i+1}, \dots, n_{k-1}, n_k; v) \quad (3). \quad [4]$$

Poichè è naturale supporre che il numero dei campioni formati con certe modalità non sia maggiore del numero dei campioni formati con quelle modalità più altre, la [4] è sottoposta alla condizione che sia $f_{ik}(v) \leq f_{rt}(v)$, quando $r \leq i \leq k \leq t$.

Il caso più frequente è che la funzione [4] si riduca a dipendere dalle due sole variabili $n_i + \dots + n_k$ e v .

Dalla [4] segue che il numero dei campioni di v unità, costituenti U_v , è dato da $U_v = f_{om}(v) = f(n_0, n_1, \dots, n_m; v)$.

È evidente che la funzione $f_{ik}(v) = f(n_i, n_{i+1}, \dots, n_{k-1}, n_k; v)$ varia a seconda della operazione campionaria prescelta.

Nel nostro lavoro ci limiteremo a considerare, come casi particolari, due operazioni campionarie, che coincidono per $v = 1$: *l'estrazione a sorte con ripetizione o bernoulliana e l'estrazione casuale in blocco*. I due universi generati dalle due operazioni suddette sono denominati rispettivamente *bernoulliano* e *in blocco*.

Identificate le unità della popolazione con le palline di una urna, il procedimento di *estrazione con ripetizione* consiste nello scegliere ognuna delle v unità del campione tra le N unità di P , indipendentemente dal risultato delle estrazioni precedenti e perchè ciò possa avvenire è indispensabile che ciascuna unità estratta venga rimessa nella popolazione prima della estrazione successiva. In tal modo il numero delle estrazioni possibili è illimitato e quindi può anche essere $v > N$. È immediato allora che, nello universo dei campioni, il numero dei campioni diversi, di v unità, formati con le sole modalità $x_i, x_{i+1}, \dots, x_{k-1}, x_k$ è dato dal numero delle disposizioni con ripetizione di $(n_i + n_{i+1} + \dots + n_{k-1} + n_k)$ oggetti di classe v , e quindi la *funzione di struttura* ha l'espressione

$$f_{ik}(v) = (n_i + n_{i+1} + \dots + n_{k-1} + n_k)^v. \quad [5]$$

L'universo è quindi costituito da N^v campioni.

(*) Conveniamo di dare significato anche all'espressione $f_{ik}(0)$ che dà la frequenza (assoluta) del campione privo di elementi, o vuoto, definendo $f(n_i, n_{i+1}, \dots, n_k; 0) = 1$ in quanto vi è una sola possibilità di non scegliere alcun elemento della popolazione.

Il procedimento di *estrazione in blocco* consiste invece nell'estrarre v delle N unità della popolazione, non una ad una, ma v alla volta, considerando uguali due campioni se costituiti dalle stesse unità, indipendentemente dall'ordine con cui queste si presentano. Deve essere perciò $v \leq N$. È immediato che in questo caso la *funzione di struttura* ha la forma

$$f_{ik}(v) = \begin{cases} 0 & \text{per } n_i + \dots + n_k < v \\ \binom{n_i + n_{i+1} + \dots + n_{k-1} + n_k}{v} & \text{per } n_i + \dots + n_k \geq v \end{cases} \quad [6]$$

ossia, nell'universo dei campioni di v unità, il numero dei campioni formati con le sole modalità $x_i, x_{i+1}, \dots, x_{k-1}, x_k$ è dato dal numero delle combinazioni di $(n_i + n_{i+1} + \dots + n_{k-1} + n_k)$ oggetti a v a v . Perciò l'universo è costituito da $\binom{N}{v}$ campioni.

DISTRIBUZIONE DELL'ESTREMO SUPERIORE E DELL'ESTREMO INFERIORE NELL'UNIVERSO \mathbf{U}_v .

4. — Determinare la distribuzione dell'estremo superiore significa ricercare il numero delle volte in cui la generica modalità x_k (per $k = 0, 1, 2, \dots, m$) si presenta come estremo superiore nei campioni che costituiscono l'universo \mathbf{U}_v . Tale numero, che indichiamo con $s_k(v)$, è dato dalla differenza tra il numero di tutti i campioni, di v unità, formati con le modalità non superiori ad x_k e il numero dei campioni, dello stesso numero di unità, formati con le modalità inferiori ad x_k ossia

$$s_k(v) = f(n_0, n_1, \dots, n_k; v) - f(n_0, n_1, \dots, n_{k-1}; v) \quad [7]$$

È immediato che $\sum_0^m s_k(v) = f(n_0, n_1, \dots, n_m; v)$ e quindi le frequenze relative hanno l'espressione:

$$\bar{s}_k(v) = \frac{f(n_0, \dots, n_{k-1}, n_k; v) - f(n_0, \dots, n_{k-1}; v)}{f(n_0, \dots, n_m; v)} \quad [8]$$

e la funzione di ripartizione, sommando le [8] da 0 a k , è data da

$$S_k(v) = \frac{f(n_0, \dots, n_k; v)}{f(n_0, \dots, n_m; v)}. \quad [9]$$

Il risultato [9] è evidente, perchè il valore S_k della funzione di ripartizione non è altro, in questo caso, che il rapporto, rispetto a $f(n_0, n_1, \dots, n_m; v)$, del numero dei campioni che hanno per estremo superiore modalità non superiori ad x_k , cioè del numero di tutti i campioni formati con le modalità x_0, x_1, \dots, x_k .

Osserviamo che si ha ovviamente $\bar{s}_k = S_k - S_{k-1}$.

Analogamente per determinare la distribuzione dell'estremo inferiore nell'universo U , bisogna ricercare il numero delle volte in cui la generica modalità x_k (per $k = 0, 1, 2, \dots, m$) si presenta come estremo inferiore nei campioni che costituiscono lo universo. Tale frequenza $i_k(v)$ non è altro che la differenza tra il numero di tutti i campioni formati con le modalità non inferiori ad x_k e il numero di tutti i campioni formati con le modalità superiori ad x_k e cioè

$$i_k(v) = f(n_k, n_{k+1}, \dots, n_m; v) - f(n_{k+1}, n_{k+2}, \dots, n_m; v), \quad [10]$$

mentre la frequenza relativa, essendo $\sum_0^m i_k(v) = f(n_0, \dots, n_m; v)$, è

$$\bar{i}_k(v) = \frac{f(n_k, n_{k+1}, \dots, n_m; v) - f(n_{k+1}, n_{k+2}, \dots, n_m; v)}{f(n_0, \dots, n_m; v)} \quad [11]$$

e, sommando le [11] da 0 a k , la funzione di ripartizione è data da

$$\begin{aligned} I_k(v) &= \frac{f(n_0, \dots, n_m; v) - f(n_{k+1}, \dots, n_m; v)}{f(n_0, \dots, n_m; v)} = \\ &= 1 - \frac{f(n_{k+1}, \dots, n_m; v)}{f(n_0, \dots, n_m; v)}. \end{aligned} \quad [12]$$

Notiamo che si ha ovviamente $i_k = I_k - I_{k-1}$.

DISTRIBUZIONE DEL CAMPO DI VARIAZIONE
NELL'UNIVERSO U ,

5. — Si fissino due modalità x_r e x_t (con $r \leq t$) e si determini il numero dei campioni, di v unità, che hanno contemporaneamente x_r come estremo inferiore e x_t come estremo superiore. Esso è evidentemente dato dalla differenza tra il numero dei campioni di v unità che hanno estremo superiore x_t e le altre modalità non inferiori ad x_r , e il numero dei campioni che hanno come estremo superiore x_t e le altre modalità maggiori di x_r , ossia

$$\begin{aligned} s_{rt} &= [f(n_r, \dots, n_t; v) - f(n_r, \dots, n_{t-1}; v)] - \\ &\quad - [f(n_{r+1}, \dots, n_t; v) - f(n_{r+1}, \dots, n_{t-1}; v)]. \end{aligned} \quad [13]$$

Il numero cercato può anche determinarsi come differenza tra il numero dei campioni, di v unità, che hanno estremo inferiore x_r e modalità non superiori ad x_t e il numero dei campioni che hanno estremo inferiore x_r e modalità minori di x_t , ossia

$$\begin{aligned} i_{rt} &= [f(n_r, \dots, n_t; v) - f(n_{r+1}, \dots, n_t; v)] - \\ &\quad - [f(n_r, \dots, n_{t-1}; v) - f(n_{r+1}, \dots, n_{t-1}; v)] \end{aligned}$$

e tale formula è evidentemente identica alla [13].

Considerato ora un valore w del campo di variazione è immediato che la frequenza relativa di w si ottiene, dividendo per $f(n_0, n_1, \dots, n_m; v)$, la somma dei valori s_{rt} per tutti i valori di r e di t degli indici tali che $x_t - x_r = w$. Indicando con $\varphi(w)$ tale frequenza, la legge di distribuzione del campo di variazione ha perciò l'espressione:

$$\begin{aligned} \varphi(w) &= \frac{1}{f(n_0, n_1, \dots, n_m; v)} \sum_{\substack{rt \\ x_t - x_r = w}} [f(n_r, \dots, n_t; v) - f(n_{r+1}, \dots, n_{t-1}; v) - \\ &\quad - f(n_{r+1}, \dots, n_t; v) + f(n_{r+1}, \dots, n_{t-1}; v)]. \end{aligned} \quad [14]$$

Nel caso particolare che le modalità siano in progressione aritmetica, ossia $x_u = x_0 + u d$ ($u = 0, 1, \dots, m$), w deve essere

multiplo di d e quindi della forma $w_q = q \bar{d}$ ($q = 0, 1, \dots, m$), ed in tal caso la [14] diviene

$$\varphi(q \bar{d}) = \frac{1}{f(n_0, n_1, \dots, n_m; \nu)} \sum_0^{m-q} [f(n_h, \dots, n_{h+q}; \nu) - f(n_h, \dots, n_{h+q-1}; \nu) - f(n_{h+1}, \dots, n_{h+q}; \nu) + f(n_{h+1}, \dots, n_{h+q-1}; \nu)]. \quad [15]$$

MEDIA ARITMETICA DEGLI ESTREMI E DEL CAMPO DI VARIAZIONE

6. — Se vale l'ipotesi enunciata alla fine del n. 2 per la definizione di media aritmetica, è possibile calcolare le medie aritmetiche $\bar{x}^{(s)}(\nu) = \sum_0^m x_k \bar{s}_k(\nu)$ degli estremi superiori dei campioni

dell'universo U_ν , e $\bar{x}^{(i)}(\nu) = \sum_0^m x_k \bar{i}_k(\nu)$ degli estremi inferiori.

Aggiungendo l'ipotesi che fra le modalità si possa definire anche la sottrazione, con le proprietà formali di cui gode nell'aritmetica, si ha, per la [8],

$$\bar{x}^{(s)}(\nu) = \sum_0^m x_k \frac{f(n_0, \dots, n_k; \nu) - f(n_0, \dots, n_{k-1}; \nu)}{f(n_0, n_1, \dots, n_m; \nu)} \quad [16]$$

e, con facili calcoli,

$$\bar{x}^{(s)}(\nu) = x_m - \sum_0^{m-1} (x_{k+1} - x_k) \frac{f(n_0, \dots, n_k; \nu)}{f(n_0, \dots, n_m; \nu)} \quad [17]$$

od anche, ricordando la [9],

$$\bar{x}^{(s)}(\nu) = x_m - \sum_0^{m-1} (x_{k+1} - x_k) S_k(\nu). \quad [18]$$

Analogamente, per la distribuzione dell'estremo inferiore, si ha la media aritmetica

$$\bar{x}^{(i)}(\nu) = x_0 + \sum_0^{m-1} (x_{k+1} - x_k) \frac{f(n_{k+1}, \dots, n_m; \nu)}{f(n_0, \dots, n_m; \nu)}, \quad [19]$$

ossia, per la [12],

$$x^{(i)}(v) = x_0 + \sum_0^{m-1} (x_{k+1} - x_k) [I - I_k(v)].$$

Se la popolazione è a modalità equispaziata, le [17] e le [19] divengono rispettivamente

$$\bar{x}^{(s)}(v) = x_m - d \sum_0^{m-1} \frac{f(n_0, \dots, n_k; v)}{f(n_0, \dots, n_m; v)} \quad [20]$$

e

$$\bar{x}^{(i)}(v) = x_0 + d \sum_0^{m-1} \frac{f(n_{k+1}, \dots, n_m; v)}{f(n_0, \dots, n_m; v)}. \quad [21]$$

7. — La media aritmetica del campo di variazione è

$$\bar{w} = \bar{x}^{(s)} - \bar{x}^{(i)} \quad (4) \quad [22]$$

perchè la variabile statistica « campo di variazione » è la differenza delle due variabili statistiche « estremo superiore » ed « estremo inferiore » ed il valore medio della differenza di due variabili statistiche è uguale alla differenza dei valori medi delle variabili stesse.

La [22] equivale alla seguente :

$$\bar{w} = (x_m - x_0) - \sum_0^{m-1} (x_{k+1} - x_k) \frac{f(n_0, \dots, n_k; v) - f(n_{k+1}, \dots, n_m; v)}{f(n_0, \dots, n_m; v)} \quad [23]$$

(4) È facile dimostrare direttamente la [22]. Infatti ricordando le [4] e le [14] si ha :

$$\begin{aligned} \bar{w} &= \sum w \varphi(w) = \frac{1}{f_{om}} \sum_{r \leq t} (f_{rt} - f_{r,t-1} - f_{r+1,t} + f_{r+1,t-1}) (x_t - x_r) = \\ &= \sum_{t=0}^m x_t \frac{\left(\sum_0^t f_{rt} - \sum_1^t f_{rt} \right) - \left(\sum_0^t f_{r,t-1} - \sum_1^t f_{r,t-1} \right)}{f_{om}} = \\ &= \frac{1}{f_{om}} \left[\sum_t \sum_r x_r (f_{rt} - f_{r+1,t}) - \sum_t \sum_r x_r (f_{rt} - f_{r+1,t-1}) \right] = \\ &= \sum_{t=0}^m x_t \frac{f_{0t} - f_{0,t-1}}{f_{om}} - \sum_r x_r \frac{f_{rm} - f_{r+1,m}}{f_{om}} = \bar{x}^{(s)} - \bar{x}^{(i)}. \end{aligned}$$

od anche alla

$$\bar{w} = (x_m - x_0) - \sum_0^{m-1} (x_{k+1} - x_k) (1 - S_k - I_k).$$

Se poi la popolazione è equispaziata, la [23] diviene

$$\bar{w} = d \left[m - \frac{\sum_0^{m-1} f(n_0, \dots, n_k; v) + f(n_{k+1}, \dots, n_m; v)}{f(n_0, \dots, n_m; v)} \right].$$

CASO PARTICOLARE DELL'UNIVERSO BERNOULLIANO.

8. — Sono state già ottenute le distribuzioni degli estremi e del campo di variazione nell'universo bernoulliano, ma, per quel che ci consta, chi le ha determinate è partito da formule piuttosto complesse per giungere, mediante semplificazioni di calcolo, alle formule alle quali noi perveniamo particolareggiando le formule dei paragrafi precedenti, da noi ottenute immediatamente ⁽⁵⁾.

Per quanto detto al n. 3, nell'universo bernoulliano si ha $f_{ik}(v) = (n_i + n_{i+1} + \dots + n_k)^v$ e quindi le [7], [8] [9], forniscono per l'estremo superiore le seguenti formule (il cui significato è ben precisato da quanto precede):

$$s'_k(v) = (n_0 + \dots + n_k)^v - (n_0 + \dots + n_{k-1})^v \quad [25]$$

$$\bar{s}'_k(v) = \frac{(n_0 + \dots + n_k)^v - (n_0 + \dots + n_{k-1})^v}{N^v} \quad [26]$$

$$S'_k(v) = \frac{(n_0 + \dots + n_k)^v}{N^v} \quad [27]$$

⁽⁵⁾ GUMBEL E.J.: *Statistical Theory of Extreme Values and Some Practical Applications*. «Nat. Bur. of Standards. Appl. Math. Series.» 33. 1954.

BURR I.: *Calculation of exact sampling distribution of ranges from a discrete population*. «Ann. Math. Stat.» vol. 26, n° 3, settembre 1955.

od anche, introducendo le frequenze relative delle modalità della popolazione $\bar{n}_i = \frac{n_i}{N}$,

$$\bar{s}'_k(v) = (\bar{n}_0 + \dots + \bar{n}_k)^v - (\bar{n}_0 + \dots + \bar{n}_{k-1})^v \quad [28]$$

$$S'_k(v) = (\bar{n}_0 + \dots + \bar{n}_k)^v. \quad [29]$$

Le [10], [11], [12] forniscono per l'estremo inferiore

$$i'_k(v) = (n_k + \dots + n_m)^v - (n_{k+1} + \dots + n_m)^v \quad [30]$$

$$\begin{aligned} \bar{i}'_k(v) &= \frac{(n_k + \dots + n_m)^v - (n_{k+1} + \dots + n_m)^v}{N^v} = \\ &= (\bar{n}_k + \dots + \bar{n}_m)^v - (\bar{n}_{k+1} + \dots + \bar{n}_m)^v \end{aligned} \quad [31]$$

$$I'_k(v) = 1 - \frac{(n_{k+1} + \dots + n_m)^v}{N^v} = 1 - (\bar{n}_{k+1} + \dots + \bar{n}_m)^v \quad [32]$$

e le [14] danno la legge di distribuzione del campo di variazione :

$$\begin{aligned} \varphi'(w) &= \frac{1}{N^v} \sum_{\substack{r,t \\ x_t - x_r = w \\ r \leq t}} [(n_r + \dots + n_t)^v - (n_r + \dots + n_{t-1})^v - \\ &\quad - (n_{r+1} + \dots + n_t)^v + (n_{r+1} + \dots + n_{t-1})^v] = \\ &= \sum_{\substack{r,t \\ x_t - x_r = w \\ r \leq t}} [(\bar{n}_r + \dots + \bar{n}_t)^v - (\bar{n}_r + \dots + \bar{n}_{t-1})^v - \\ &\quad - (\bar{n}_{r+1} + \dots + \bar{n}_t)^v + (\bar{n}_{r+1} + \dots + \bar{n}_{t-1})^v] \end{aligned} \quad [33]$$

9. — La media aritmetica degli estremi superiori è, per la [17]

$$\bar{x}'^{(s)}(v) = x_m - \sum_0^{m-1} (x_{k-1} - x_k) (\bar{n}_0 + \dots + \bar{n}_k)^v \quad [34]$$

e, nel caso delle popolazioni equispaziate,

$$\bar{x}'^{(s)}(v) = x_m - d \sum_0^{m-1} (\bar{n}_0 + \dots + \bar{n}_k)^v, \quad [34']$$

mentre quella degli estremi inferiori è, per la [19],

$$\bar{x}'^{(i)}(v) = x_0 + \sum_0^{m-1} (x_{k+1} - x_k) (\bar{n}_{k+1} + \dots + \bar{n}_m)^v \quad [35]$$

e, nel caso delle popolazioni equispaziate,

$$\bar{x}'^{(i)}(v) = x_0 + d \sum_0^{m-1} (\bar{n}_{k+1} + \dots + \bar{n}_m)^v.$$

È immediato che, se la popolazione è costituita da più di una modalità effettiva valgono le disuguaglianze

$$x_0 < \bar{x}'^{(s)}(v) < x_m \quad [36]$$

$$x_0 < \bar{x}'^{(i)}(v) < x_m, \quad [37]$$

che mettono in evidenza che la media aritmetica degli estremi superiori (inferiori) è minore (maggiore) dell'estremo superiore (inferiore) della popolazione. Perciò, se si usano l'estremo superiore e quello inferiore del campione per stimare rispettivamente l'estremo superiore e l'estremo inferiore della popolazione, tali stime sono sempre *distorte*.

La [23] fornisce per la media aritmetica del campo di variazione

$$\begin{aligned} \bar{w}' &= x_m - x_0 - \sum_0^{m-1} (x_{k+1} - x_k) \cdot \\ &\cdot [(\bar{n}_0 + \dots + \bar{n}_k)^k + (\bar{n}_{k+1} + \dots + \bar{n}_m)^v], \end{aligned} \quad [38]$$

la [24], per la popolazione equispaziata,

$$\bar{w}' = d \left\{ m - \sum_0^{m-1} [(\bar{n}_0 + \dots + \bar{n}_k)^v + (\bar{n}_{k+1} + \dots + \bar{n}_m)^v] \right\}$$

e per quella che è anche simmetrica, ossia per la quale si ha

$$n_k = n_{m-k},$$

$$\bar{w}' = d \left[m - 2 \sum_0^{m-1} (\bar{n}_0 + \dots + \bar{n}_k)^v \right]. \quad [39]$$

Dalla [38] segue poi la disuguaglianza

$$\overline{w}' \leq x_m - x_0, \quad [40]$$

nella quale vale il segno di uguaglianza nel solo caso $m = 0$, tolto il quale, la stima del campo di variazione della popolazione, effettuata mediante il campo di variazione del campione, è *distorta*.

UNIVERSO BERNOULLIANO: VARIAZIONI AL VARIARE DEL NUMERO DELLE UNITÀ DEL CAMPIONE

10. — Dimostriamo le seguenti proprietà:

I) *La frequenza relativa dell'estremo superiore della popolazione, come estremo superiore dei campioni, ossia $\overline{s}'_m(v)$, cresce al crescere di v .*

Escludiamo il caso $n_m = 0$ perchè in tal caso è come se la modalità x_m non esistesse per la popolazione, e il discorso varrebbe per $\overline{s}_{m-1}(v)$.

Dalla [28] si ha

$$\overline{s}'_m(v) = 1 - (\overline{n}_0 + \dots + \overline{n}_{m-1})^v$$

e poichè $\overline{n}_0 + \dots + \overline{n}_{m-1} < \overline{n}_0 + \dots + \overline{n}_m = 1$, al crescere del numero delle unità del campione, $(\overline{n}_0 + \dots + \overline{n}_{m-1})^v$ decresce e quindi $\overline{s}'_m(v)$ aumenta, ossia sono più frequenti i campioni che hanno per estremo superiore l'estremo superiore della popolazione.

Nello stesso modo si dimostra che:

II) *La frequenza relativa del primo valore x_0 come estremo inferiore dei campioni ossia $i'_0(v)$, cresce al crescere di v .*

III) *Qualunque sia h , la funzione di ripartizione $S'_h(v)$, $[I'_h(v)]$, dell'estremo superiore [inferiore] decresce [cresce] al crescere di v .*

Da quest'ultima proprietà tenendo presente la [18] segue immediatamente che

IV) *La media aritmetica degli estremi superiori cresce al crescere di v e, poichè si è visto alla formula [36] che $\bar{x}^{(s)}(v) < x_m$, si avvicina all'estremo superiore della popolazione.*

V) *La media aritmetica degli estremi inferiori decresce al crescere di v , e poichè si è visto alla formula (37) che $\bar{x}^{(i)} v > x_0$, si avvicina all'estremo inferiore della popolazione.*

Dalla [22] e dalle due proprietà precedenti segue la proprietà:

VI) *la media aritmetica del campo di variazione cresce al crescere di v , e si avvicina quindi al campo di variazione della popolazione.*

II. — Aumentando il numero delle unità dei campioni, il teorema I afferma che sono più frequenti i campioni aventi per estremo superiore l'estremo superiore della popolazione, mentre il teorema II attesta che sono più frequenti i campioni aventi per estremo inferiore l'estremo inferiore della popolazione. Il teorema seguente afferma in più che, *con l'aumentare del numero delle unità del campione, sono più frequenti i campioni che hanno per estremi inferiore e superiore gli estremi inferiore e superiore della popolazione, ossia che sono più frequenti i campioni che hanno per campo di variazione il campo di variazione della popolazione.*

Escludiamo il caso banale che gli elementi della popolazione siano raggruppati in una sola classe x_0 , perchè in tal caso ogni campione, qualunque sia v , può essere costituito solo da elementi di quella classe e quindi ha, come la popolazione, campo di variazione zero. Ammettiamo allora anzitutto $n_0 \neq 0$ ed $n_m \neq 0$. In tal modo $x_m - x_0$ è il campo di variazione della popolazione ed anche il valore massimo della distribuzione del campo di variazione nell'universo dei campioni. Indicando con $\varphi'_m(v)$ la frequenza relativa di quest'ultimo, per la [33] si ha

$$\varphi'_m(v) = 1 - (1 - \bar{n}_0)^v - (1 - \bar{n}_m)^v + (1 - \bar{n}_0 - \bar{n}_m)^v, \quad [41]$$

ed inoltre

$$\varphi'_m(v+1) = 1 - (1 - \bar{n}_0)^{v+1} - (1 - \bar{n}_m)^{v+1} + (1 - \bar{n}_0 - \bar{n}_m)^{v+1}, \quad [42]$$

e la differenza

$$\begin{aligned}\varphi'_m(v+1) - \varphi'_m(v) &= \bar{n}_0 [(1 - \bar{n}_0)^v - (1 - \bar{n}_0 - \bar{n}_m)^v] + \\ &+ \bar{n}_m [(1 - \bar{n}_m)^v - (1 - \bar{n}_0 - \bar{n}_m)^v]\end{aligned}$$

è evidentemente positiva.

UNIVERSO BERNOULLIANO : LE DISTRIBUZIONI DEGLI ESTREMI PER POPOLAZIONI A DISTRIBUZIONE NON CRESCENTE O NON DECRESCENTE

12. — Escludiamo il caso $v = 1$, nel quale si ha $n_k = s'_k = i'_k$ e quindi la non decrescenza, o la non crescenza, della distribuzione della popolazione comporta, rispettivamente, la non decrescenza, o la non crescenza, delle distribuzioni degli estremi. Escludiamo anche il caso $n_0 = 0$. Per $v > 1$ dimostriamo che, se la distribuzione della popolazione non è decrescente, ossia è tale che $n_{k+1} \geq n_k$, ed $n_0 > 0$, la distribuzione dell'estremo superiore è crescente.

Infatti

$$s'_0 = n_0^v; s'_1 = (n_0 + n_1)^v - n_0^v = n_1^v + \binom{v}{1} n_1^{v-1} n_0 + \dots + \binom{v}{v-1} n_1 n_0^{v-1}$$

e quindi se $n_1 \geq n_0$, il primo addendo è non inferiore ad n_0^v e quindi $s'_1 > s'_0$.

Si ha inoltre, ricordando l'identità

$$A^v - B^v = (A - B) (A^{v-1} + A^{v-2} B + \dots + A B^{v-2} + B^{v-1}),$$

$$\begin{aligned}s'_k &= (n_0 + \dots + n_{k-1} + n_k)^v - (n_0 + \dots + n_{k-1})^v = \\ &= n_k [(n_0 + \dots + n_k)^{v-1} + \\ &+ (n_0 + \dots + n_k)^{v-2} (n_0 + \dots + n_{k-1}) + \dots + (n_0 + \dots + n_{k-1})^{v-1}]\end{aligned}$$

ed analogamente

$$\begin{aligned}s'_{k+1} &= n_{k+1} [(n_0 + \dots + n_{k+1})^{v-1} + \\ &+ (n_0 + \dots + n_{k+1})^{v-2} (n_0 + \dots + n_k) + \dots + (n_0 + \dots + n_k)^{v-1}].\end{aligned}$$

Se $n_{k+1} \geq n_k$ ($n_k > 0$ perchè $n_0 > 0$) poichè la somma fra parentesi quadre di s'_{k+1} è maggiore di quella fra parentesi quadre di s'_k , si ha $s'_{k+1} > s'_k$.

Perciò il teorema è completamente dimostrato. Ne segue immediatamente che *per una popolazione con distribuzione delle frequenze non decrescente e con $n_0 > 0$, la successione delle frequenze relative dell'estremo superiore è crescente.*

Analogamente si dimostra che *per una popolazione con distribuzione delle frequenze non crescente, ossia tale che $n_k \leq n_{k+1}$, e con $n_m > 0$, la distribuzione dell'estremo inferiore e la corrispondente successione delle frequenze relative è decrescente.*

CASO PARTICOLARE DELL'UNIVERSO IN BLOCCO.

13. — Poichè la funzione di struttura dell'universo in blocco è

$$f_{ik}(v) = \begin{cases} 0 & \text{per } n_i + \dots + n_k < v \\ \left(n_i + n_{i+1} + \dots + n_{k-1} + n_k \right) & \text{per } n_i + \dots + n_k \geq v, \end{cases} \quad [6]$$

indicando con j ⁽⁶⁾ il più piccolo degli indici k per i quali $n_0 + \dots + n_k \geq v$, dalla [7] si ha, per la frequenza assoluta dei campioni che, nell'universo, hanno x_k come estremo superiore,

$$s'_k(v) = \begin{cases} 0 & \text{per } k < j \\ \left(n_0 + \dots + n_j \right) & \text{per } k = j \\ \left(n_0 + \dots + n_k \right) - \left(n_0 + \dots + n_{k-1} \right) & \text{per } k > j, \end{cases} \quad [43]$$

od anche, introducendo la funzione di ripartizione della popolazione

$$F_k = n_0 + n_1 + \dots + n_k, \quad [44]$$

(6) o talvolta con $j(v)$ per indicare la dipendenza di j da v .

$$s_k''(\nu) = \begin{cases} 0 & \text{per } k < j \\ \binom{F_j}{\nu} & \text{per } k = j \\ \binom{F_k}{\nu} - \binom{F_{k-1}}{\nu} & \text{per } k > j. \end{cases} \quad [45]$$

Dividendo le [45] per $\sum_{k=0}^m s_k''(\nu) = \binom{N}{\nu}$ si ha la legge di distribuzione dell'estremo superiore :

$$\bar{s}_k''(\nu) = \begin{cases} 0 & \text{per } k < j \\ \frac{\binom{F_j}{\nu}}{\binom{N}{\nu}} & \text{per } k = j \\ \frac{\binom{F_k}{\nu} - \binom{F_{k-1}}{\nu}}{\binom{N}{\nu}} & \text{per } k > j \end{cases} \quad [46]$$

mentre la funzione di ripartizione ha, per la [9], l'espressione

$$S_k''(\nu) = \sum_0^k \bar{s}_k''(\nu) = \begin{cases} 0 & \text{per } k < j \\ \frac{\binom{F_k}{\nu}}{\binom{N}{\nu}} & \text{per } k \geq j. \end{cases} \quad [47]$$

Analogamente a quanto è stato fatto per la distribuzione dello estremo superiore, se si indica con l ^(?), il più grande degli indici h per i quali $n_h + n_{h+1} + \dots + n_{m-1} + n_m \geq \nu$, le [6] e le [10]

^(?) o talvolta con $l(\nu)$ per mettere in evidenza la dipendenza di l da ν .

forniscono per la frequenza assoluta i_k'' dell'estremo inferiore, nell'universo dei campioni, l'espressione

$$i_k''(v) = \begin{cases} \binom{n_k + \dots + n_m}{v} - \binom{n_{k+1} + \dots + n_m}{v} & \text{per } k < l \\ \binom{n_l + \dots + n_m}{v} & \text{per } k = l \\ 0 & \text{per } k > l \end{cases} \quad [48]$$

ossia, introducendo le [44],

$$i_k''(v) = \begin{cases} \binom{N - F_{k-1}}{v} - \binom{N - F_k}{v} & \text{per } k < l \\ \binom{N - F_{l-1}}{v} & \text{per } k = l \\ 0 & \text{per } k > l. \end{cases} \quad [49]$$

Il quoziente di $i_k''(v)$ per $\sum_{k=0}^m i_k''(v) = \binom{N}{v}$ dà la legge di distribuzione dell'estremo inferiore

$$\bar{i}_k''(v) = \begin{cases} \frac{\binom{N - F_{k-1}}{v} - \binom{N - F_k}{v}}{\binom{N}{v}} & \text{per } k < l \\ \frac{\binom{N - F_{l-1}}{v}}{\binom{N}{v}} & \text{per } k = l \\ 0 & \text{per } k > l. \end{cases} \quad [50]$$

Inoltre la [12] fornisce per la funzione di ripartizione l'espressione

$$I_k''(v) = \begin{cases} 1 - \frac{\binom{N - F_k}{v}}{\binom{N}{v}} & \text{per } k < l \\ 0 & \text{per } k \geq l. \end{cases} \quad [51]$$

Notiamo che quando la popolazione è simmetrica, e ciò avviene se, per ogni k , $n_k = n_{m-k}$, si ha $i_k'' = s_{m-k}'', \bar{i}_k'' = \bar{s}_{m-k}'', I_k'' = S_{m-k}''$ ed ovviamente $s_k'' = i_{m-k}'', \bar{s}_k'' = \bar{i}_{m-k}'', S_k'' = I_{m-k}''$.

14. — Vale il seguente:

TEOREMA: La frequenza relativa dell'estremo superiore della popolazione, come estremo superiore dei campioni nell'universo campionario, se $n_m > 0$ e $v > 1$, è maggiore nell'universo il blocco che nell'universo bernoulliano ⁽⁸⁾.

Si ha infatti nell'universo in blocco per $v \leq N - n_m$:

$$\begin{aligned} \bar{s}_m'' &= 1 - \frac{\binom{N-n_m}{v}}{\binom{N}{v}} = 1 - \frac{(N-n_m)(N-n_m-1)\dots(N-n_m-v+1)}{N \cdot (N-1)\dots(N-v+1)} = \\ &= 1 - \left(1 - \frac{n_m}{N}\right) \left(1 - \frac{n_m}{N-1}\right) \dots \left(1 - \frac{n_m}{N-v+1}\right) > 1 - \left(1 - \frac{n_m}{N}\right)^v = \bar{s}_m'. \end{aligned}$$

Se $v > N - n_m$, si ha $\bar{s}_m'' = 1$, $\bar{s}_m' = 1 - \left(1 - \frac{n_m}{N}\right)^v$, e quindi il teorema è completamente dimostrato.

Analogamente vale il

TEOREMA: La frequenza relativa dell'estremo inferiore della popolazione, come estremo inferiore dei campioni nell'universo campionario, se $n_0 > 0$ e $v > 1$, è minore nell'universo in blocco che nell'universo bernoulliano.

15. — Determiniamo ora la distribuzione del campo di variazione nell'universo in blocco. Con i simboli già introdotti le [13] forniscono

$$\begin{aligned} s_{r,t}'' &= \binom{n_r + \dots + n_t}{v} - \binom{n_r + \dots + n_{t-1}}{v} - \\ &\quad - \binom{n_{r+1} + \dots + n_t}{v} + \binom{n_{r+1} + \dots + n_{t-1}}{v} \end{aligned} \quad [52]$$

⁽⁸⁾ Tale proposizione, da me enunciata inizialmente per $n_m = 1$, mi è stata suggerita, nell'attuale formulazione più generale dal prof. A. HERZEL.

(dove si considerano nulli tutti i coefficienti binomiali che hanno i termini superiori minori di v), ovvero

$$s''_{rt} = \binom{F_t - F_{r-1}}{v} - \binom{F_{t-1} - F_{r-1}}{v} - \binom{F_t - F_r}{v} + \binom{F_{t-1} - F_{r-1}}{v}$$

con la convenzione di considerare $F_{-1} = 0$

Le [14] forniscono per il campo di variazione la legge di distribuzione:

$$\begin{aligned} \varphi''(w) = \frac{1}{\binom{N}{v}} \sum_{\substack{r,t \\ x_t - x_r = w \\ r \leq t}} & \left[\binom{F_t - F_{r-1}}{v} - \binom{F_{t-1} - F_{r-1}}{v} - \right. \\ & \left. - \binom{F_t - F_r}{v} + \binom{F_{t-1} - F_r}{v} \right]. \end{aligned} \quad [53]$$

Dalla [52] deriva che, se $n_r = 0$ o $n_t = 0$, $s''_{rt} = 0$, ossia il contributo alla [53] di modalità, aventi frequenza zero, è nullo. Perciò non si toglie nulla alla generalità della trattazione se si considera una popolazione equispaziata, ossia a modalità in progressione aritmetica, in quanto se la popolazione non è così fatta, lo diviene immediatamente, interpolandovi modalità di frequenza zero.

Con le notazioni precedenti la [15] o la [53] forniscono la frequenza relativa ⁽⁹⁾

$$\begin{aligned} \varphi''(q, d) = \frac{1}{\binom{N}{v}} \sum_h^{m-q} & \left[\binom{F_{h+q} - F_{h-1}}{v} - \binom{F_{h+q-1} - F_{h-1}}{v} - \right. \\ & \left. - \binom{F_{h+q} - F_h}{v} + \binom{F_{h+q-1} - F_h}{v} \right] \end{aligned} \quad [54]$$

⁽⁹⁾ Se \bar{n} è il massimo delle n_h , $\varphi''(q, d) = 0$ per ogni q per il quale $\bar{n} - q < v$.

Svolgendo la [54] si ottiene

$$\begin{aligned} \varphi''(q, d) = & \frac{1}{\binom{N}{v}} \left\{ \sum_h^{m-q} \binom{F_{h+q} - F_{h-1}}{v} - 2 \sum_h^{m-(q-1)} \binom{F_{h+(q-1)} - F_h}{v} + \right. \\ & + \sum_h^{m-(q-2)} \binom{F_{h+(q-2)} - F_{h-1}}{v} - \left[\binom{F_{q-2}}{v} + \binom{N - F_{m-q+1}}{v} \right] + \\ & \left. + \left[\binom{F_{q-1}}{v} + \binom{N - F_{m-q}}{v} \right] \right\}. \end{aligned} \quad [55]$$

Le [55] mettono in evidenza che vi è una certa ricorrenza per il calcolo di $\varphi''(q, d)$, cosicchè dopo che sono stati determinati $\varphi''(0)$, $\varphi''(d)$, $\varphi''(2d)$... $\varphi''((q-1)d)$, la determinazione di $\varphi''(q, d)$ comporta solo il calcolo del primo e dello ultimo termine della [55]. Inoltre le stesse [55] permettono di calcolare facilmente la funzione di ripartizione $\Phi''(w_q) = \Phi''(q, d)$: infatti, sommandole, si ricava

$$\begin{aligned} \Phi''(q, d) = & \frac{1}{\binom{N}{v}} \left[\sum_h^{m-q} \binom{F_{h+q} - F_{h-1}}{v} - \sum_h^{m-(q-1)} \binom{F_{h+(q-1)} - F_{h-1}}{v} + \right. \\ & \left. + \binom{F_{q-1}}{v} + \binom{N - F_{m-q}}{v} \right] \end{aligned}$$

ossia

$$\Phi''(q, d) = \frac{1}{\binom{N}{v}} \left[\sum_h^{m-q} \binom{n_h + \dots + n_{h+q}}{v} - \sum_h^{m-q} \binom{n_h + \dots + n_{h+q-1}}{v} \right]. \quad [56]$$

Osserviamo che si ha ovviamente

$$\sum_0^m \varphi''(q, d) = \frac{\binom{F_m}{v}}{\binom{N}{v}} = 1.$$

16. — Per la media aritmetica, degli estremi superiori si ha

$$\bar{x}''^{(s)}(v) = x_m - \sum_i^{m-1} (x_{k+1} - x_k) \frac{\binom{F_k}{v}}{\binom{N}{v}} \quad [57]$$

e, nel caso che la popolazione sia equispaziata,

$$\bar{x}''^{(s)} = x_m - d \sum_j^{m-1} \frac{\binom{F_k}{v}}{\binom{N}{v}}; \quad [58]$$

mentre per la media aritmetica degli estremi inferiori

$$\bar{x}''^{(i)}(v) = x_0 + \sum_0^{l-1} (x_{k+1} - x_k) \frac{\binom{N - F_k}{v}}{\binom{N}{v}} \quad [59]$$

che, nel caso che la popolazione sia equispaziata, diviene

$$\bar{x}''^{(i)} = x_0 + d \sum_0^{l-1} \frac{\binom{N - F_k}{v}}{\binom{N}{v}}. \quad [60]$$

Per la media aritmetica del campo di variazione si ha

$$\bar{w}'' = (x_m - x_0) - \sum_j^{m-1} (x_{k+1} - x_k) \frac{\binom{F_k}{v}}{\binom{N}{v}} - \sum_0^{l-1} (x_{k+1} - x_k) \frac{\binom{N - F_k}{v}}{\binom{N}{v}}, \quad [61]$$

che, se la popolazione è equispaziata, si trasforma nella

$$\bar{w}'' = d \left[m - \sum_j^{m-1} \frac{\binom{F_k}{v}}{\binom{N}{v}} - \sum_0^{l-1} \frac{\binom{N - F_k}{v}}{\binom{N}{v}} \right] \quad [62]$$

e, se è anche simmetrica, nella

$$\bar{w}'' = d \left[m - 2 \sum_j^{m-1} \frac{\binom{F_k}{v}}{\binom{N}{v}} \right].$$

Come si vede dalla [57] se $v > F_{m-1}$ il sottraendo è nullo e perciò $\bar{x}''^{(s)} = x_m$; in ogni altro caso $\bar{x}''^{(s)} < x_m$ e quindi in generale

$$\bar{x}''^{(s)} \leq x_m \quad [63]$$

Analogamente si vede che

$$\bar{x}''^{(i)} \geq x_0, \quad [64]$$

dove il segno di uguaglianza vale per $v > N - F_0$. Cosicchè, per v più grande del maggiore fra F_{m-1} e $N - F_0$, è $\bar{w}'' = x_m - x_0$; altrimenti $\bar{w}'' < x_m - x_0$ e quindi in generale è $\bar{w}'' \leq x_m - x_0$. Perciò la stima di x_m , effettuata mediante l'estremo superiore del campione, è *corretta* per $v > F_{m-1}$, è *distorta* nel caso opposto; la stima di x_0 , fatta mediante l'estremo inferiore del campione è *corretta* per $v > N - F_0$ e *distorta* nel caso opposto; ed infine la stima di $x_m - x_0$, effettuata col campo di variazione del campione è *corretta* per v maggiore sia di F_{m-1} che di $N - F_0$, ed è *distorta* in ogni altro caso.

UNIVERSO IN BLOCCO : VARIAZIONI AL VARIARE DEL NUMERO DELLE UNITÀ DEL CAMPIONE

17. — Dimostriamo che la *frequenza relativa dell'estremo superiore della popolazione come estremo superiore nell'universo dei campioni* ossia $\bar{s}_m''(v)$, cresce al crescere di v fino ad 1, rimanendo poi costantemente uguale ad 1.

Si può escludere anzitutto il caso $n_m = 0$ perchè in tal caso è come se la modalità x_m non esistesse per la popolazione: sia perciò $F_m = N$ e $F_{m-1} < N$.

Per $1 \leq v < F_{m-1}$ si ha

$$\bar{s}_m''(v) = 1 - \frac{\binom{F_{m-1}}{v}}{\binom{N}{v}}; \bar{s}_m''(v+1) = 1 - \frac{\binom{F_{m-1}}{v+1}}{\binom{N}{v+1}} = 1 - \frac{\binom{F_{m-1}}{v}}{\binom{N}{v}} \cdot \frac{F_{m-1}-v}{N-v}$$

e poichè $F_{m-1} < N$, è $F_{m-1} - v < N - v$ e quindi

$$\bar{s}_m''(v+1) > \bar{s}_m''(v).$$

Per $v = F_{m-1}$ si ha

$$\bar{s}_m''(v) = 1 - \frac{1}{\binom{N}{v}}, \quad \bar{s}_m''(v+1) = 1$$

e quindi

$$\bar{s}_m''(v+1) > \bar{s}_m''(v).$$

Per $v > F_{m-1}$ si ha

$$\bar{s}_m''(v) = \bar{s}_m''(v+1) = \dots = \bar{s}_m''(N) = 1$$

e in tal caso ogni campione ha lo stesso estremo superiore della popolazione.

Si è visto così che al crescere del numero delle unità del campione sono più frequenti i campioni che hanno per estremo superiore l'estremo superiore della popolazione.

Analogamente si dimostra che la frequenza relativa del primo valore x_0 , come estremo inferiore ossia $\bar{i}_0(v)$, cresce, al crescere di v , fino ad 1, rimanendo poi costantemente uguale ad 1, ossia al crescere del numero delle unità del campione sono più frequenti i campioni che hanno come estremo inferiore l'estremo inferiore della popolazione.

Fissiamo ora la modalità x_h ; per tutti i valori di v compresi nell'intervallo aperto a sinistra $F_{h-1} \rightarrow F_h$, ossia per ogni v tale che $F_{h-1} < v \leq F_h$, si ha

$$\bar{s}_h'(v) = \frac{\binom{F_h}{v}}{\binom{N}{v}}.$$

Studiamo come varia in tal caso $\bar{s}_h''(v)$ al variare di v . Dimostriamo che, fissata una modalità, $x_h < x_m$ la frequenza di questa come primo valore, diverso da zero, della legge di distribuzione $s_h''(v)$ decresce al crescere di v .

Sia per esempio

$$F_{h-1} < v < v+1 < \dots < v+a \leq F_h;$$

per $p < a$ si ha

$$\bar{s}_h''(v+p) = \frac{\binom{F_h}{v+p}}{\binom{N}{v+p}} = \frac{F_h! (N-v-p)!}{N! (F_h-v-p)!} = \frac{F_h! (N-v-p-1)!}{N! (F_h-v-p-1)!} \cdot \frac{N-v-p}{F_h-v-p} = \bar{s}_h''(v+p+1) \frac{N-v}{F_h-v} > \bar{s}_h''(v+p+1)$$

essendo $N > F_h$; per $p = a$ si ha

$$\bar{s}_h''(v+p) = \frac{\binom{F_h}{v+p}}{\binom{N}{v+p}} \quad \text{e} \quad \bar{s}_h''(v+p+1) = 0,$$

e quindi in tutti e due casi si ha

$$\bar{s}_h''(v+p) > \bar{s}_h''(v+p+1).$$

Nello stesso modo si può dimostrare che, fissata una modalità $x_h > x_0$, la frequenza di questa come ultimo valore, diverso da zero della legge di distribuzione $\bar{i}_h''(v)$ decresce al crescere di v .

Dimostriamo anche che per h qualunque la funzione di ripartizione S_h' dell'estremo superiore è non crescente al crescere del numero delle unità del campione.

Per $h = m$, è $S_h'' = 1$ per ogni valore di v (che può essere al massimo uguale ad N).

Se $h < m$ e se $v < v+1 \leq F_h$ si ha

$$\begin{aligned} S_h''(v+1) &= \frac{\binom{F_h}{v+1}}{\binom{N}{v+1}} = \frac{F_h! (N-v-1)!}{N! (F_h-v-1)!} = \frac{F_h! (N-v)!}{N! (F_h-v)!} \frac{F_h-v}{N-v} = \\ &= \frac{\binom{F_h}{v}}{\binom{N}{v}} \cdot \frac{F_h-v}{N-v} = S_h''(v) \frac{F_h-v}{N-v} < S_h''(v), \end{aligned}$$

essendo $F_k < N$; se invece $v = F_k < v + 1$, si ha

$$S_h''(v) = \frac{\binom{F_k}{v}}{\binom{N}{v}} \quad \text{ed} \quad S_h''(v+1) = 0$$

e quindi

$$S_h''(v) > S_h''(v+1);$$

se poi $v > F_k$, sia $S_h''(v)$ sia $S_h''(v+1)$ sono nulli e quindi in generale

$$S_h''(v) \geq S_h''(v+1) \quad \text{c.v.d.}$$

Analogamente si dimostra che per k qualunque la funzione di ripartizione I_k'' dell'estremo inferiore è non decrescente al crescere di v .

18. — Passiamo alla media aritmetica degli estremi superiori: escluso il caso $v > F_{m-1}$, nel quale, come si è visto al n. 17, ogni campione ha come estremo superiore l'estremo superiore della popolazione, dimostriamo che la media aritmetica degli estremi superiori, cresce al crescere di v da 1 a F_{m-1} , ossia che, aumentando il numero delle unità del campione, la media aritmetica si avvicina all'estremo superiore della popolazione.

Supponiamo $n_m > 0$, perchè se $n_m = 0$ ed $n_{m-1} > 0$ tanto varrebbe considerare la popolazione x_0, \dots, x_{m-1} .

Se $v < v+1 \leq F_{m-1}$ si ha

$$\begin{aligned} \bar{x}''^{(s)}(v) &= x_m - \sum_{j(v)}^{m-1} (x_{k+1} - x_k) \frac{\binom{F_k}{v}}{\binom{N}{v}} = \\ &= x_m - \sum_{j(v)}^{m-1} (x_{k+1} - x_k) \frac{F_k! (N-v)!}{N! (F_k-v)!} \\ \bar{x}''^{(s)}(v+1) &= x_m - \sum_{j(v+1)}^{m-1} (x_{k+1} - x_k) \frac{\binom{F_k}{v+1}}{\binom{N}{v+1}} = \\ &= x_m - \sum_{j(v+1)}^{m-1} (x_{k+1} - x_k) \frac{\binom{F_k}{v}}{\binom{N}{v}} \cdot \frac{F_k-v}{N-v} \end{aligned}$$

e quindi, poichè la sommatoria che compare in $\bar{x}''^{(s)}(v+1)$ ha un numero di termini minore od uguale a quello della sommatoria che è in $\bar{x}''^{(s)}(v)$, essendo $j(v) \leq j(v+1)$, ed inoltre ogni addendo di $\bar{x}''^{(s)}(v+1)$ è minore del corrispondente che compare in $\bar{x}''^{(s)}(v)$, perchè $F_k - v < N - v$, si ha $\bar{x}''^{(s)}(v+1) < \bar{x}''^{(s)}(v)$.

Se invece $v = F_{m-1} < v+1$

$$\bar{x}''^{(s)}(v) = x_m - (x_m - x_{m-1}) \frac{\binom{F_{m-1}}{v}}{\binom{N}{v}} < x_m$$

ed $\bar{x}''^{(s)}(v+1) = x_m$ e quindi anche

$$\bar{x}''^{(s)}(v+1) < \bar{x}''^{(s)}(v) \quad \text{c. v. d.}$$

In complesso si ha, per v crescente, che la media aritmetica degli estremi superiori non decresce al crescere del numero delle unità del campione, ossia si avvicina all'estremo superiore della popolazione.

Analogamente si dimostra che la media aritmetica degli estremi inferiori.

$$\bar{x}''^{(i)}(v) = x_0 + \sum_0^{l-1} (x_{k+1} - x_k) \frac{\binom{N - F_k}{v}}{\binom{N}{v}}$$

non cresce al crescere del numero v delle unità del campione, anzi decresce per v crescente nell'intervallo $1 \leq v \leq (N - F_0)$, mentre è costantemente uguale a x_0 per $v > N - F_0$, cosicchè al crescere del numero delle unità del campione la media aritmetica suddetta si avvicina all'estremo inferiore della popolazione.

19. — Passando alla variazione della frequenza del valore massimo del campo di variazione dei campioni (valore massimo che è uguale al campo di variazione della popolazione) supponiamo, come al n. 11, che la popolazione sia costituita da più elementi, distribuiti in più di una classe, con $n_0 \neq 0$ ed $n_m \neq 0$ ed indichiamo con $\phi_m''(v)$ la frequenza relativa di tale valore massimo.

Dalla [54] si ha

$$\varphi_m''(\nu) = 1 - \frac{\binom{N-n_0}{\nu}}{\binom{N}{\nu}} - \frac{\binom{N-n_m}{\nu}}{\binom{N}{\nu}} + \frac{\binom{N-n_0-n_m}{\nu}}{\binom{N}{\nu}} \quad [65]$$

e quindi, nell'ipotesi $\nu + 1 < N$,

$$\varphi_m''(\nu + 1) = 1 - \frac{\binom{N-n_0}{\nu+1}}{\binom{N}{\nu+1}} - \frac{\binom{N-n_m}{\nu+1}}{\binom{N}{\nu+1}} + \frac{\binom{N-n_0-n_m}{\nu+1}}{\binom{N}{\nu+1}} \quad [66]$$

Osserviamo ora che se $N - a > \nu$

$$\frac{\binom{N-a}{\nu}}{\binom{N}{\nu}} - \frac{\binom{N-a}{\nu+1}}{\binom{N}{\nu+1}} = \frac{\binom{N-a}{\nu}}{\binom{N}{\nu}} - \frac{\binom{N-a}{\nu}}{\binom{N}{\nu}}. \quad [67]$$

$$\cdot \frac{N-a-\nu}{N-\nu} = \frac{\binom{N-a}{\nu}}{\binom{N}{\nu}} \cdot \frac{a}{N-\nu} > 0.$$

e se $N - a = \nu$

$$\frac{\binom{N-a}{\nu}}{\binom{N}{\nu}} - \frac{\binom{N-a}{\nu+1}}{\binom{N}{\nu+1}} = \frac{\binom{N-a}{\nu}}{\binom{N}{\nu}} > 0 \quad [68]$$

Si può ora enunciare il

TEOREMA; *La frequenza relativa del massimo valore della distribuzione del campo di variazione cresce, per ν crescente e non maggiore del massimo fra $N - n_m$ e $N - n_0$, per il quale, come per ogni altro valore di ν maggiore di questo numero e non superiore ad N , vale 1.*

Dimostriamo infatti che la differenza

$$\begin{aligned} \varphi_m''(v+1) - \varphi_m''(v) = & \left[\frac{\binom{N-n_0}{v}}{\binom{N}{v}} - \frac{\binom{N-n_0}{v+1}}{\binom{N}{v+1}} \right] + \\ & + \left[\frac{\binom{N-n_m}{v}}{\binom{N}{v}} - \frac{\binom{N-n_m}{v+1}}{\binom{N}{v+1}} \right] - \\ & - \left[\frac{\binom{N-n_0-n_m}{v}}{\binom{N}{v}} - \frac{\binom{N-n_0-n_m}{v+1}}{\binom{N}{v+1}} \right] \end{aligned} \quad [69]$$

è zero, se $v > N - n_0$ e $v > N - n_m$, altrimenti è sempre positiva.

Infatti se $N - n_0 < v$ e $N - n_m < v$ la [65] e la [66] danno

$$\varphi_m''(v) = 1, \quad \varphi_m''(v+1) = 1 \quad \text{e quindi} \quad \varphi_m''(v+1) - \varphi_m''(v) = 0$$

Se invece $N - n_0 \geq v$ e $N - n_m < v$, nel secondo membro della [69] rimane solamente il primo addendo fra parentesi quadre, che per la [67] o per la [68] è positivo. Data la simmetria, avviene lo stesso se $N - n_m \geq v$ e $N - n_0 < v$.

Se $N - n \geq 0$ e $N - n_m \geq 0$ si possono presentare i tre casi:

I) $N - n_0 - n_m < v$; II) $v \leq N - n_0 - n_m < v + 1$ ossia $N - n_0 - n_m = v$; III) $N - n_0 - n_m \geq v + 1$, che esaminiamo separatamente.

I) Se $N - n_0 - n_m < v$, il terzo addendo fra parentesi quadre è nullo e i primi due sono positivi per la [67] e per la [68] e quindi

$$\varphi_m''(v+1) - \varphi_m''(v) > 0$$

II) Se $N - n_0 - n_m = v$ le [69] divengono

$$\varphi_m''(v+1) - \varphi_m''(v) = \frac{\binom{N-n_0}{v}}{\binom{N}{v}} \frac{n_0}{N-v} + \frac{\binom{N-n_m}{v}}{\binom{N}{v}} \frac{n_m}{N-v} -$$

$$-\frac{1}{\binom{N}{v}} = \frac{1}{\binom{N}{v}} \left[\binom{N-n_0}{v} \frac{n_0}{N-v} + \binom{N-n_m}{v} \frac{n_m}{N-v} - 1 \right]$$

e poichè $N - n_0 - n_m = v$, ne segue $N - n_0 > v$, $N - n_m > v$ e quindi

$$\begin{aligned} \frac{1}{\binom{N}{v}} \left[\binom{N-n_0}{v} \frac{n_0}{N-v} + \binom{N-n_m}{v} \frac{n_m}{N-v} - 1 \right] &> \\ &> \frac{1}{\binom{N}{v}} \frac{n_0 + n_m - N + v}{N-v} = 0 \end{aligned}$$

III) se $N - n_0 - n_m \geq v + 1$ le [69] divengono, per le [67],

$$\begin{aligned} \varphi_m''(v+1) - \varphi_m''(v) &= \frac{\binom{N-n_0}{v}}{\binom{N}{v}} \frac{n_0}{N-v} + \frac{\binom{N-n_m}{v}}{\binom{N}{v}} \frac{n_m}{N-v} - \\ &- \frac{\binom{N-n_0-n_m}{v}}{\binom{N}{v}} \frac{n_0+n_m}{N-v} = \left[\frac{\binom{N-n_0}{v}}{\binom{N}{v}} - \frac{\binom{N-n_0-n_m}{v}}{\binom{N}{v}} \right] \cdot \\ &\cdot \frac{n_0}{N-v} + \left[\frac{\binom{N-n_m}{v}}{\binom{N}{v}} - \frac{\binom{N-n_0-n_m}{v}}{\binom{N}{v}} \right] \frac{n_m}{N-v} > 0 \end{aligned}$$

perchè, essendo $n_0 > 0$ e $n_m > 0$, è

$$\binom{N-n_0}{v} > \binom{N-n_0-n_m}{v} \quad \text{e} \quad \binom{N-n_m}{v} > \binom{N-n_0-n_m}{v},$$

e così il teorema è completamente dimostrato.

Si è dimostrato dunque che, *al crescere del numero delle unità del campione, sono sempre più frequenti i campioni il cui campo di variazione coincide con il campo di variazione della popolazione.*

Per la variazione della media aritmetica del campo di variazione osserviamo che poichè $\bar{x}^{(s)}(v)$ è non decrescente, al crescere di v , e $\bar{x}^{(t)}(v)$ è non crescente, la loro differenza e quindi, per la (22), la media della distribuzione del campo di variazione è non decrescente al crescere di v . Si ha perciò che tale media si approssima al campo di variazione della popolazione quando aumenta il numero delle unità del campione.

UNIVERSO IN BLOCCO : LE DISTRIBUZIONI DEGLI ESTREMI PER POPOLAZIONI A DISTRIBUZIONE NON CRESCENTE O NON DECRESCENTE.

20. — Escluso il caso $v = 1$ nel quale $n_k = \bar{s}_k'$, è bene distinguere i due casi, $k = j(v)$ e $k > j(v)$.

I caso: $k = j(v)$. Osserviamo anzitutto che il rapporto $\frac{(c+a+1)(c+a+2)\dots(c+2a-1)(c+2a)}{(a-b+1)(a-b+2)\dots(2a-b-1)(2a-b)}$ è non inferiore a 2, per a, b, c interi non negativi e $0 < b < a$.

Infatti per $b = 1$ e $c = 0$ esso diviene

$$\frac{(a+1)\dots(2a-1)2a}{a(a+1)\dots(2a-1)} = \frac{2a}{a} = 2$$

e per $b > 1$ e $c \geq 0$

$$\frac{(c+a+1)\dots(c+2a-1)(c+2a)}{(a-b+1)\dots(2a-b)} > \frac{(a+1)(a+2)\dots 2a}{a(a+1)\dots(2a-1)} = 2$$

Riassumendo si ha dunque

$$\left. \begin{aligned} &> 2 \quad \text{per } c > 0 \\ &> 2 \quad \text{per } c = 0 \quad \text{e } b > 1 \\ &= 2 \quad \text{per } c = 0 \quad \text{e } b = 1 \end{aligned} \right\}$$

Dall'osservazione ora fatta segue il

TEOREMA: Se $n_i \leq n_{i+1}$ e $v > 1$ allora $\bar{s}_i' < \bar{s}_{i+1}'$.

Per quanto detto in precedenza deve essere $F_{i-1} < v$ e si potrà perciò porre $v = F_{i-1} + b$ con b intero positivo minore di n_i .

Consideriamo dapprima il caso $n_j = n_{j+1}$: si ha

$$\binom{F_j}{\nu} = \binom{F_{j-1} + n_j}{F_{j-1} + b} = \frac{(F_{j-1} + b + 1) \dots (F_{j-1} + n_j)}{(n_j - b)!}$$

$$\binom{F_{j+1}}{\nu} = \binom{F_{j-1} + 2 n_j}{F_{j-1} + b} = \frac{(F_{j-1} + b + 1) \dots (F_{j-1} + 2 n_j)}{(2 n_j - b)!}$$

e quindi

$$\frac{\bar{s}_{j+1}'''}{\bar{s}_j'''} = \frac{\binom{F_{j+1}}{\nu} - \binom{F_j}{\nu}}{\binom{F_j}{\nu}} = \frac{\binom{F_{j+1}}{\nu}}{\binom{F_j}{\nu}} - 1 = \frac{(F_{j-1} + n_j + 1) \dots (F_{j-1} + 2 n_j)}{(n_j - b + 1) \dots (2 n_j - b)} - 1.$$

Per l'osservazione fatta si ha

$$\frac{(F_{j-1} + n_j + 1) \dots (F_{j-1} + 2 n_j)}{(n_j - b + 1) \dots (2 n_j - b)} > 2$$

a meno che non sia $F_{j-1} = 0$ e $b = 1$, ma in questo caso si ha $\nu = 1$ e ciò è stato escluso.

Si ha quindi $\frac{\bar{s}_{j+1}'''}{\bar{s}_j'''} > 1$ e perciò $\bar{s}_j''' < \bar{s}_{j+1}'''$.

Se $n_j < n_{j+1}$

$$\bar{s}_{j+1}''' = \binom{F_{j-1} + n_j + n_{j+1}}{\nu} - \binom{F_{j-1} + n_j}{\nu} > \binom{F_{j-1} + 2 n_j}{\nu} -$$

$$- \binom{F_{j-1} + n_j}{\nu} \geq \bar{s}_j'''$$

e con ciò si è dimostrato l'asserto.

II caso: $k > j(\nu)$. Premettiamo il seguente

LEMMA: *Dati i due numeri interi e positivi a e b , con $a - b \geq \nu - 1$, l'espressione*

$$\binom{a+b}{\nu} - 2 \binom{a}{\nu} + \binom{a-b}{\nu} \quad [70]$$

è maggiore di zero.

Dimostriamo il lemma per induzione :

per $b = 1$ si ha ⁽¹⁰⁾, nell'ipotesi $a - 1 \geq v > 1$ (necessaria perchè tutti i simboli scritti abbiano significato) :

$$\begin{aligned} \binom{a+1}{v} - 2 \binom{a}{v} + \binom{a-1}{v} &= \binom{a}{v-1} + \binom{a}{v} - \\ - 2 \binom{a}{v} + \binom{a}{v} - \binom{a-1}{v-1} &= \binom{a}{v-1} - \binom{a-1}{v-1} \end{aligned}$$

dove l'ultimo membro è positivo perchè $a > a - 1$.

Per $b = 2$ si ha, nell'ipotesi $a - 2 \geq v > 1$, e quindi $a - 1 > v$, (il che fa sì che i simboli scritti abbiano significato) :

$$\begin{aligned} \binom{a+2}{v} - 2 \binom{a}{v} + \binom{a-2}{v} &= \binom{a+1}{v-1} + \binom{a+1}{v} - 2 \binom{a}{v} + \binom{a-1}{v} - \\ - \binom{a-2}{v-1} &= \left[\binom{a+1}{v} - 2 \binom{a}{v} + \binom{a-1}{v} \right] + \left[\binom{a+1}{v-1} - \binom{a-2}{v-1} \right] \end{aligned}$$

espressione che è positiva perchè $a + 1 > a - 2$, e quindi $\binom{a+1}{v-1} - \binom{a-2}{v-1} > 0$, e perchè come si è visto precedentemente, il primo termine fra parentesi quadrate è positivo.

In generale se $a - b \geq v > 1$ si ha

$$\begin{aligned} \binom{a+b}{v} - 2 \binom{a}{v} + \binom{a-b}{v} &= \binom{a+b-1}{v-1} + \binom{a+b-1}{v} - 2 \binom{a}{v} + \\ + \binom{a-b+1}{v} - \binom{a-b}{v-1} &= \left[\binom{a+(b-1)}{v} - 2 \binom{a}{v} + \binom{a-(b-1)}{v} \right] + \\ + \left[\binom{a+b-1}{v-1} - \binom{a-b}{v-1} \right] \end{aligned}$$

e poichè il secondo termine tra parentesi quadre è positivo essendo $a + b - 1 > a - b$, l'espressione è maggiore di zero se lo è

⁽¹⁰⁾ Usando la formula $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$ o, ciò che è lo stesso,

$$\binom{n-1}{k} = \binom{n}{k} - \binom{n-1}{k-1}$$

il primo termine fra parentesi quadrate, che non è altro che ciò che si ottiene dalla [70], quando si sostituisce $b - 1$ in luogo di b .

Il lemma è quindi dimostrato.

Ne segue il

TEOREMA : Se $n_k \leq n_{k+1}$ per $k > j$, allora $\bar{s}_k' < \bar{s}_{k+1}'$.

Si ha infatti $F_{k-1} = F_k - n_k$, con $F_{k-1} \geq v$ essendo $k > j$, $F_{k+1} = F_k + n_{k+1}$ e perciò

$$\bar{s}_k' = \binom{F_k}{v} - \binom{F_k - n_k}{v} \text{ e } \bar{s}_{k+1}' = \binom{F_k + n_{k+1}}{v} - \binom{F_k}{v}$$

da cui

$$\begin{aligned} \bar{s}_{k+1}' - \bar{s}_k' &= \binom{F_k + n_{k+1}}{v} - 2 \binom{F_k}{v} + \\ &+ \binom{F_k - n_k}{v} \geq \binom{F_k + n_k}{v} - 2 \binom{F_k}{v} + \binom{F_k - n_k}{v} \end{aligned}$$

e quindi, essendo $F_k - n_k \geq v > 1$, per il lemma precedente si ha

$$\bar{s}_{k+1}' - \bar{s}_k' > 0 \text{ e perciò } \bar{s}_k' < \bar{s}_{k+1}'.$$

Da quanto dimostrato segue che : fissato il numero $v > 1$ delle unità di ciascun campione, se la distribuzione delle frequenze di una popolazione è da $j(v)$ in poi non decrescente, la successione delle frequenze dell'estremo superiore è crescente da \bar{s}_j' in poi.

In particolare si ha : per una popolazione, con distribuzione delle frequenze non decrescente, la legge di distribuzione dell'estremo superiore è non decrescente per ogni v .

Analogamente si ha che per una popolazione, con distribuzione delle frequenze non crescente, le successioni delle frequenze, assolute e relative, dell'estremo inferiore sono non crescenti per ogni v .

Nel caso della popolazione a modalità numeriche equispaziate, le [34'] forniscono per $v = 1, 2, \dots, 8$ le seguenti medie aritmetiche:

$$\begin{aligned}\bar{x}'^{(s)}(1) &= x_0 + 1,5 d = x_8 - 1,5 d \\ \bar{x}'^{(s)}(2) &= x_0 + 1,968 d = x_8 - 1,032 d \\ \bar{x}'^{(s)}(3) &= x_0 + 2,203 d = x_8 - 0,797 d \\ \bar{x}'^{(s)}(4) &= x_0 + 2,352 d = x_8 - 0,648 d \\ \bar{x}'^{(s)}(5) &= x_0 + 2,456 d = x_8 - 0,544 d \\ \bar{x}'^{(s)}(6) &= x_0 + 2,536 d = x_8 - 0,464 d \\ \bar{x}'^{(s)}(7) &= x_0 + 2,600 d = x_8 - 0,400 d \\ \bar{x}'^{(s)}(8) &= x_0 + 2,653 d = x_8 - 0,347 d.\end{aligned}$$

Consideriamo poi la popolazione di 16 elementi distribuiti binomialmente secondo le cinque modalità x_0, x_1, x_2, x_3, x_4 con le frequenze 1; 4; 6; 4; 1, e l'universo dei campioni di 1, 2, ..., 16 unità generati applicando ad essa l'operazione campionaria di estrazione con ripetizione. Mediante le [28] si ottengono per ciascun universo le distribuzioni riportate nella tab. 2.

Le [34'] forniscono le medie aritmetiche:

$$\begin{aligned}\bar{x}'^{(s)}(1) &= x_0 + 2 d = x_4 - 2 d \\ \bar{x}'^{(s)}(2) &= x_0 + 2,547 d = x_4 - 1,453 d \\ \bar{x}'^{(s)}(3) &= x_0 + 2,820 d = x_4 - 1,180 d \\ \bar{x}'^{(s)}(4) &= x_0 + 2,995 d = x_4 - 1,005 d \\ \bar{x}'^{(s)}(5) &= x_0 + 3,119 d = x_4 - 0,881 d \\ \bar{x}'^{(s)}(6) &= x_0 + 3,215 d = x_4 - 0,785 d \\ \bar{x}'^{(s)}(7) &= x_0 + 3,291 d = x_4 - 0,709 d \\ \bar{x}'^{(s)}(8) &= x_0 + 3,353 d = x_4 - 0,647 d \\ \bar{x}'^{(s)}(9) &= x_0 + 3,406 d = x_4 - 0,594 d \\ \bar{x}'^{(s)}(10) &= x_0 + 3,452 d = x_4 - 0,548 d \\ \bar{x}'^{(s)}(11) &= x_0 + 3,492 d = x_4 - 0,508 d \\ \bar{x}'^{(s)}(12) &= x_0 + 3,528 d = x_4 - 0,472 d \\ \bar{x}'^{(s)}(13) &= x_0 + 3,560 d = x_4 - 0,440 d \\ \bar{x}'^{(s)}(14) &= x_0 + 3,590 d = x_4 - 0,410 d \\ \bar{x}'^{(s)}(15) &= x_0 + 3,617 d = x_4 - 0,383 d \\ \bar{x}'^{(s)}(16) &= x_0 + 3,641 d = x_4 - 0,359 d.\end{aligned}$$

TAB. 2. — Leggi di distribuzione dell'estremo superiore, negli universi dei campioni, di 1, 2, ..., 16 unità, generati operando bernoullianamente su una popolazione binomiale di 16 elementi.

$x \backslash v$	1 (*)	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
x_0	0,0625	0,004	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
x_1	0,2500	0,094	0,030	0,009	0,003	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
x_2	0,3750	0,375	0,295	0,214	0,151	0,105	0,072	0,050	0,034	0,024	0,016	0,011	0,008	0,005	0,004	0,002
x_3	0,2500	0,406	0,499	0,550	0,571	0,573	0,564	0,547	0,525	0,501	0,476	0,450	0,424	0,400	0,376	0,354
x_4	0,0625	0,121	0,176	0,227	0,275	0,321	0,364	0,403	0,441	0,475	0,508	0,539	0,568	0,595	0,620	0,644
Totale	1,0000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

(*) Per mettere in evidenza la binomialità della distribuzione, le frequenze della prima colonna sono state riportate senza l'approssimazione della terza cifra decimale, al contrario di quanto è stato fatto per le altre frequenze;

Nel caso dell'operazione campionaria di estrazione in blocco, si hanno per gli universi dei campioni di $v = 1, 2, \dots, 8$ unità, generati nella popolazione binomiale di otto elementi, le distribuzioni riportate nella tab. 3.

TAB. 3. — *Leggi di distribuzione dell'estremo superiore negli universi dei campioni di 1, 2, ..., 8 unità generati operando in blocco su una popolazione binomiale di 8 elementi.*

v	1	2	3	4	5	6	7	8
$x x_0$	0,125	0	0	0	0	0	0	0
x_1	0,375	0,214	0,071	0,014	0	0	0	0
x_2	0,375	0,536	0,554	0,486	0,375	0,250	0,125	0
x_3	0,125	0,250	0,375	0,500	0,625	0,750	0,875	1
Totale.	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1

Nel caso delle popolazioni equispaziate le [58] forniscono, per ciascuna distribuzione, le medie aritmetiche :

$$\begin{aligned} \bar{x}''^{(s)}(1) &= x_0 + 1,5 d = x_3 - 1,5 d \\ \bar{x}''^{(s)}(2) &= x_0 + 2,036 d = x_3 - 0,964 d \\ \bar{x}''^{(s)}(3) &= x_0 + 2,304 d = x_3 - 0,696 d \\ \bar{x}''^{(s)}(4) &= x_0 + 2,486 d = x_3 - 0,514 d \\ \bar{x}''^{(s)}(5) &= x_0 + 2,625 d = x_3 - 0,375 d \\ \bar{x}''^{(s)}(6) &= x_0 + 2,750 d = x_3 - 0,250 d \\ \bar{x}''^{(s)}(7) &= x_0 + 2,875 d = x_3 - 0,125 d \\ \bar{x}''^{(s)}(8) &= x_0 + 3 d = x_3. \end{aligned}$$

Per la popolazione binomiale di 16 elementi si hanno le distribuzioni riportate nella tab. 4.

TAB. 4. — Leggi di distribuzione dell'estremo superiore negli universi dei campioni, di 1, 2, 3, ..., 16 unità generati operando in blocco su una popolazione binomiale di 16 elementi.

$x \backslash y$	1 (*)	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
x_0	0,0625	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x_1	0,2500	0,083	0,018	0,003	0,000	0	0	0	0	0	0	0	0	0	0	0
x_2	0,3750	0,375	0,279	0,179	0,106	0,058	0,029	0,013	0,005	0,001	0,000	0	0	0	0	0
x_3	0,2500	0,417	0,516	0,568	0,582	0,567	0,534	0,487	0,432	0,374	0,312	0,250	0,187	0,125	0,062	0
x_4	0,0625	0,125	0,187	0,250	0,312	0,375	0,437	0,500	0,563	0,625	0,688	0,750	0,813	0,875	0,938	1
Totale . . .	1,0000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1

(*) Si veda la nota alla tab. 2.

Applicando le [58] si hanno le medie aritmetiche :

$$\begin{aligned}
 \bar{x}''^{(s)}(1) &= x_0 + 2d = x_4 - 2d \\
 \bar{x}''^{(s)}(2) &= x_0 + 2,584d = x_4 - 1,416d \\
 \bar{x}''^{(s)}(3) &= x_0 + 2,875d = x_4 - 1,125d \\
 \bar{x}''^{(s)}(4) &= x_0 + 3,065d = x_4 - 0,935d \\
 \bar{x}''^{(s)}(5) &= x_0 + 3,206d = x_4 - 0,794d \\
 \bar{x}''^{(s)}(6) &= x_0 + 3,317d = x_4 - 0,683d \\
 \bar{x}''^{(s)}(7) &= x_0 + 3,409d = x_4 - 0,591d \\
 \bar{x}''^{(s)}(8) &= x_0 + 3,487d = x_4 - 0,513d \\
 \bar{x}''^{(s)}(9) &= x_0 + 3,558d = x_4 - 0,442d \\
 \bar{x}''^{(s)}(10) &= x_0 + 3,624d = x_4 - 0,376d \\
 \bar{x}''^{(s)}(11) &= x_0 + 3,688d = x_4 - 0,312d \\
 \bar{x}''^{(s)}(12) &= x_0 + 3,750d = x_4 - 0,250d \\
 \bar{x}''^{(s)}(13) &= x_0 + 3,813d = x_4 - 0,187d \\
 \bar{x}''^{(s)}(14) &= x_0 + 3,875d = x_4 - 0,125d \\
 \bar{x}''^{(s)}(15) &= x_0 + 3,938d = x_4 - 0,062d \\
 \bar{x}''^{(s)}(16) &= x_0 + 4d = x_4.
 \end{aligned}$$

Dalle tabelle 1, 2, 3, 4 si osserva, come teoricamente era stato dimostrato ai n. 10 e 17, che la frequenza relativa dell'estremo superiore della popolazione, come estremo superiore negli universi bernoulliano e in blocco, cresce col crescere di v e che negli universi bernoulliani la frequenza relativa dell'estremo inferiore della popolazione decresce al crescere di v . Come si era detto al n. 14, la frequenza relativa dell'estremo superiore della popolazione, come estremo superiore nell'universo campionario, per $v > 1$, è maggiore nell'universo in blocco, che in quello bernoulliano.

Per $v = 1$ la distribuzione dell'estremo superiore, in entrambi gli universi, coincide con la popolazione di partenza e quindi è binomiale. Per $v > 1$ le distribuzioni presentano sempre una asimmetria, dovute ad uno spostamento a destra della moda, che al crescere di v o rimane ferma o si sposta verso l'estremo superiore della popolazione.

Osserviamo nelle tabelle 3 e 4, relative ad *universi in blocco* :

1) che $\bar{s}_m''\left(\frac{N}{4}\right) = \frac{1}{4}$; $\bar{s}_m''\left(\frac{N}{2}\right) = \frac{1}{2}$; $\bar{s}_m''\left(\frac{3N}{4}\right) = \frac{3}{4}$; 2) che la moda coincide con l'estremo superiore della popolazione per $v \geq \frac{N}{2}$.

La prima proprietà è diretta conseguenza della formula $\bar{s}_m''(v) = \frac{v}{N}$, che, come si dimostra facilmente, è valida quando $n_m = 1$. La seconda è un caso particolare della proprietà:

Se $n_m = 1$, la distribuzione dell'estremo superiore, negli universi in blocco, è unimodale per $v > \frac{N}{2}$ e la moda coincide con l'estremo superiore della popolazione. Per $v = \frac{N}{2}$ ciò avviene se $\frac{N}{2} < N - n_{m-1} - n_m$. Infatti, per $v = \frac{N}{2}$, è $\bar{s}_m''\left(\frac{N}{2}\right) = \frac{1}{2}$ e se $\frac{N}{2} < N - n_{m-1} - n_m$, è $0 < \bar{s}_{m-2}'' < \frac{1}{2}$ e $0 < \bar{s}_{m-1}'' < \frac{1}{2}$ e quindi x_m è l'unica moda. Per $v > \frac{N}{2}$, per quanto dimostrato all'inizio del n. 17, è $\bar{s}_m''(v) > \bar{s}_m''\left(\frac{N}{2}\right) = \frac{1}{2}$ e perciò anche in questo caso x_m è l'unica moda.

Per gli universi bernoulliani delle tabelle 1 e 2 la moda coincide invece con l'estremo superiore, nel primo per $v \geq 5$ e nel secondo per $v \geq 11$, e con frequenze relative crescenti, all'aumentare di v , molto meno rapidamente che nell'universo in blocco.

Si è visto dunque che l'estremo superiore della popolazione ha frequenza relativa maggiore (per ogni v maggiore di 1) nello universo in blocco che in quello bernoulliano e che, negli esempi fatti (ed è da opinare che ciò accada sempre), l'estremo superiore diventa valore più frequente prima nell'universo senza ripetizione che in quello bernoulliano. Quindi, se l'operazione di campionamento ha lo scopo di dare una stima dell'estremo superiore della popolazione, è più conveniente usare il campionamento in blocco che quello bernoulliano.

Per quanto riguarda la media aritmetica della distribuzione dell'estremo superiore, si era dimostrato al n. 10 che nell'universo bernoulliano la media cresce al crescere del numero delle unità del campione, mentre al n. 18, per l'universo in blocco, si era visto che certamente non decresce. Gli esempi fatti inducono a credere che anche per gli universi in blocco, ottenuti da popolazioni binomiali equispaziate, la media è non solo non decrescente, ma anzi crescente. Il valore $\bar{x}'^{(s)}(1) = \bar{x}''^{(s)}(1) = x_0 + \frac{m}{2}d$, che si calcola

immediatamente, è perciò il valore minimo della media dello estremo superiore.

Tolto il caso $v = 1$, a parità di v , si osserva che la media delle distribuzioni dell'estremo superiore è più vicina all'estremo superiore delle popolazioni nell'universo in blocco che in quello bernoulliano. Ciò porta alle stesse conclusioni, alle quali si è giunti prima, circa la scelta del metodo di campionamento.

22. — Data la simmetria della distribuzione della popolazione binomiale si ha

$$\bar{z}_k(v) = \bar{s}_{m-k}(v)$$

e perciò le distribuzioni dell'estremo inferiore per le popolazioni binomiali considerate si ricavano dalle tabelle precedenti usando l'uguaglianza scritta.

Tutte le proprietà osservate nei due paragrafi precedenti si mutano in quelle per l'estremo inferiore mutando dualmente la parola « superiore » con la parola « inferiore ».

23. — Per le stesse popolazioni del n. 21, con l'ipotesi aggiuntiva che siano equispaziate, si sono ottenute, per il campo di variazione, le distribuzioni e le medie aritmetiche, riportate nelle tabelle 5, 6, 7, 8.

TAB. 5. — *Leggi di distribuzioni e medie aritmetiche del campo di variazione negli universi dei campioni generati operando bernoullianamente su una popolazione equispaziata e binomiale di 8 elementi.*

$w \backslash v$	1	2	3	4	5	6	7	8
0.	1	0,313	0,109	0,040	0,015	0,006	0,002	0,000
\bar{d}	0	0,468	0,457	0,362	0,270	0,198	0,145	0,107
$2\bar{d}$	0	0,187	0,352	0,454	0,504	0,516	0,505	0,480
$3\bar{d}$	0	0,032	0,082	0,144	0,211	0,280	0,348	0,413
Totale.	1	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Media	0	0,938 \bar{d}	1,407 \bar{d}	1,703 \bar{d}	1,911 \bar{d}	2,071 \bar{d}	2,199 \bar{d}	2,305 \bar{d}

TAB. 6. — Leggi di distribuzione e medie aritmetiche del campo di variazione negli universi dei campioni generati operando bernoullianamente su una popolazione equispaziata e binomiale di 16 elementi.

ν w																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
0	1	0,273	0,084	0,027	0,009	0,003	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
d	0	0,437	0,381	0,270	0,178	0,115	0,073	0,047	0,029	0,018	0,011	0,007	0,004	0,003	0,002	0,001
$2d$	0	0,219	0,367	0,431	0,442	0,423	0,389	0,350	0,311	0,274	0,240	0,210	0,183	0,160	0,139	0,121
$3d$	0	0,063	0,146	0,231	0,306	0,368	0,417	0,453	0,478	0,494	0,502	0,504	0,499	0,493	0,483	0,471
$4d$	0	0,008	0,022	0,041	0,065	0,091	0,120	0,150	0,182	0,214	0,247	0,279	0,313	0,344	0,376	0,407
Totale.	1	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Media.	0	1,095 d	1,641 d	1,989 d	2,239 d	2,429 d	2,582 d	2,707 d	2,813 d	2,904 d	2,985 d	3,055 d	3,120 d	3,178 d	3,233 d	3,283 d

TAB. 7. — *Leggi di distribuzione e medie aritmetiche del campo di variazione negli universi dei campioni generati operando in blocco su una popolazione equispaziata e binomiale di 8 elementi.*

$\begin{matrix} v \\ w \end{matrix}$	1	2	3	4	5	6	7	8
0	1	0,214	0,036	0	0	0	0	0
d	0	0,536	0,428	0,243	0,107	0,036	0	0
$3d$	0	0,214	0,428	0,543	0,536	0,428	0,250	0
$4d$	0	0,036	0,108	0,214	0,357	0,536	0,750	1
Totale.	1	1,000	1,000	1,000	1,000	1,000	1,000	1
Media	0	1,062 d	1,608 d	1,972 d	2,250 d	2,502 d	2,750 d	3 d

Come è stato dimostrato, si osserva che, al crescere di v , cresce la frazione dei campioni aventi per campo di variazione quello della popolazione ed aumenta anche la media aritmetica.

Il campo di variazione della popolazione non coincide mai, negli esempi fatti, per gli universi bernoulliani, con la moda delle distribuzioni del campo di variazione, al contrario di quando accade per gli universi in blocco, per i quali la coincidenza si ha, per le popolazioni di 8, 16, 32 elementi, rispettivamente per

$$v \geq 6, v \geq 11 \text{ e } v \geq 22.$$

TAB. 8. — *Leggi di distribuzione e medie del campo di variazione negli universi dei campioni generati operando in blocco su una popolazione equispaziata e binomiale di 16 elementi.*

v	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
0	1	0,225	0,050	0,009	0,002	0,000	0	0	0	0	0	0	0	0	0	0
d	0	0,467	0,364	0,218	0,113	0,052	0,021	0,007	0,002	0,000	0	0	0	0	0	0
$2d$	0	0,233	0,397	0,455	0,440	0,386	0,316	0,246	0,181	0,127	0,084	0,050	0,025	0,008	0	0
$3d$	0	0,067	0,164	0,268	0,362	0,437	0,488	0,514	0,517	0,498	0,458	0,400	0,325	0,233	0,125	0
$4d$	0	0,008	0,025	0,050	0,083	0,125	0,175	0,233	0,300	0,375	0,458	0,550	0,650	0,759	0,875	1
Totale . . .	1	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1
Media . . .	0	1,167 d	1,750 d	2,132 d	2,411 d	2,635 d	2,817 d	2,973 d	3,115 d	3,248 d	3,375 d	3,500 d	3,625 d	3,750 d	3,875 d	4 d

UNIVERSO IN BLOCCO

CASO DELLE POPOLAZIONI EQUIDISTRIBUITE.

24. — Per popolazione equidistribuita intendiamo una popolazione le cui modalità hanno tutte la stessa frequenza n' .

Le [46] forniscono la legge di distribuzione

$$\bar{s}_k' = \begin{cases} 0 & \text{per } k < j \\ \frac{\binom{(j+1)n'}{v}}{\binom{(m+1)n'}{v}} & \text{per } k = j \\ \frac{\binom{(k+1)n'}{v} - \binom{kn'}{v}}{\binom{(m+1)n'}{v}} & \text{per } k > j \end{cases}$$

per l'estremo superiore, e le [50] la legge di distribuzione per l'estremo inferiore

$$\bar{i}_k'' = \begin{cases} \frac{\binom{(m-k+1)n'}{v} - \binom{(m-k)n'}{v}}{\binom{(m+1)n'}{v}} & \text{per } k < l \\ \frac{\binom{(m-l+1)n'}{v}}{\binom{(m+1)n'}{v}} & \text{per } k = l \\ 0 & \text{per } k > l \end{cases}$$

che, nel caso $n' = 1$, osservando che $m + 1 = N$, divengono rispettivamente

$$\bar{s}_k'' = \begin{cases} 0 & \text{per } k < v - 1 \\ \frac{\binom{k+1}{v} - \binom{k}{v}}{\binom{N}{v}} = \frac{\binom{k}{v-1}}{\binom{N}{v}} & \text{per } k \geq v - 1 \end{cases} \quad [71]$$

$$\bar{i}_k'' = \begin{cases} \frac{\binom{N-k}{v} - \binom{N-k-1}{v}}{\binom{N}{v}} = \frac{\binom{N-k-1}{v-1}}{\binom{N}{v}} & \text{per } k > N - v \\ 0 & \text{per } k \geq N - v \end{cases} \quad [72]$$

Per la media aritmetica, dalla [58], si ha, nell'ipotesi che la popolazione sia anche equispaziata,

$$\bar{x}''^{(s)} = x_m - d \frac{\sum_{v=1}^{m-1} \binom{k}{v-1}}{\binom{N}{v}} = x_m - d \frac{\binom{N}{v+1}}{\binom{N}{v}} = x_m - d \frac{N-v}{v+1} \quad [73]$$

od anche, dopo brevi calcoli,

$$\bar{x}''^{(s)} = x_0 + \frac{N-v-1}{v+1} d. \quad [74]$$

Ed analogamente

$$\bar{x}''^{(i)} = x_0 + \frac{N-v}{v+1} d = x_m - \frac{N-v-1}{v+1} d. \quad [75]$$

Sempre per la popolazione equispaziata ed equidistribuita, con $n' = 1$, è stata calcolata la varianza delle distribuzioni dello estremo inferiore e superiore e si è trovato

$$\sigma_s^2 = \sigma_i^2 = d^2 (N-v) \left[\frac{2N+1}{v+1} - \frac{N-v}{(v+1)^2} - \frac{2(N+1)}{v+2} \right]. \quad [76]$$

25. — In questo numero vengono messi a confronto gli universi dei campioni di 2, 3, 4, 5, unità generati dall'operazione campionaria di estrazione in blocco, che opera sulla popolazione costituita dai numeri 1, 2, 89, 90, ciascuno con frequenza 1, con insiemi di 1000 campioni, rispettivamente dello stesso numero di unità, estratti a caso e in blocco dalla stessa popolazione.

Questi insieme sono stati ricavati dai risultati delle estrazioni del gioco del lotto: in Italia, settimanalmente, in dieci città vengono estratti a caso, in blocco, 5 numeri fra i primi 90 numeri naturali, cosicchè, per ogni settimana, si può disporre di dieci campioni in blocco di 5 elementi.

Sono stati considerati i 520 campioni del 1957 e i 520 del 1958, ma, per semplificare i calcoli, si è preferito avere 1000 campioni, escludendo i risultati di due estrazioni ogni anno. Perchè tale esclusione dipendesse solo dal caso, per ogni anno, iniziando dalla prima settimana si sono ricercati, fra i primi estratti di ogni cinquina, i primi due numeri non superiori a 52. Per il 1957 si sono trovati 1 e 47 e perciò non si sono considerate le estrazioni della 1^a e della 47^a settimana e per il 1958 si sono trovati 33 e 14, e così si sono escluse la 33^a e la 14^a settimana.

Di ogni cinquina sono state considerate la prima coppia di numeri estratti, ottenendo così 1000 campioni di due unità, la prima terna, ottenendo 1000 campioni di tre unità, e così via fino a 5. Si è preso in considerazione per ogni campione, l'estremo superiore e le distribuzioni trovate sono riportate rispettivamente nella II e nella IV colonna, delle tabb. 9, 10, 11, 12. Nelle stesse tabelle sono riportate anche le distribuzioni dell'estremo superiore negli universi, distribuzioni che, tenendo conto della [71], sono date dalla legge

$$\bar{s}_k' = \begin{cases} 0 & \text{per } 1 \leq k < v - 1 \\ \frac{\binom{k-1}{v-1}}{\binom{90}{v}} & \text{per } k \geq v - 1 \end{cases}$$

Dal confronto fra i dati della tabella si può dedurre che vi è buona aderenza fra le coppie di distribuzioni. Per analizzare

TAB. 9. — *Frequenze dell'estremo superiore nell'insieme dei campioni, costituiti dai primi due estratti di 1000 cinquine del lotto, e nel corrispondente universo in blocco.*

ESTREMO SUPERIORE	FRE- QUENZE NELL'IN- SIEME DEI 1000 CAM- PIONI	FREQUENZE NELL'UNIVERSO DEI CAMPIONI		ESTREMO SUPERIORE	FRE- QUENZE NELL'IN- SIEME DEI 1000 CAM- PIONI	FREQUENZE NELL'UNIVERSO DEI CAMPIONI	
		assolute	relative a 1000 cam- pioni			assolute	relative a 1000 cam- pioni
1	—	0		52	13	51	13
2	—	1		53	21	52	13
3	—	2		54	8	53	13
4	—	3	5	55	9	54	13
5	1	4		56	12	55	14
6	3	5		57	16	56	14
7	—	6		58	12	57	14
8	3	7		59	9	58	14
9	—	8	6	60	19	59	15
10	1	9		61	14	60	15
11	1	10		62	13	61	15
12	4	11	5	63	13	62	15
13	3	12		64	16	63	16
14	4	13	6	65	11	64	16
15	6	14		66	16	65	16
16	—	15	7	67	15	66	16
17	2	16		68	13	67	17
18	7	17	8	69	14	68	17
19	4	18		70	26	69	17
20	3	19	10	71	23	70	18
21	2	20	5	72	16	71	18
22	4	21	5	73	26	72	18
23	4	22	6	74	16	73	18
24	4	23	6	75	15	74	18
25	4	24	6	76	11	75	19
26	11	25	6	77	19	76	19
27	6	26	7	78	17	77	19
28	7	27	7	79	16	78	20
29	8	28	7	80	25	79	20
30	9	29	7	81	26	80	20
31	11	30	8	82	20	81	20
32	11	31	8	83	15	82	20
33	10	32	8	84	23	83	21
34	6	33	8	85	26	84	21
35	6	34	9	86	23	85	21
36	13	35	9	87	20	86	21
37	12	36	9	88	18	87	22
38	12	37	9	89	32	88	22
39	11	38	10	90	32	89	22
40	14	39	10	<div> <div>Totale</div> <div>Media arit.</div> <div>Varianza</div> <div>Scarto quadratico medio</div> <div>Gradi di libert�</div> <div>$\chi^2 = 79,01$</div> </div>			
41	7	40	10				
42	6	41	10				
43	7	42	10				
44	8	43	11				
45	7	44	11				
46	8	45	11				
47	8	46	11				
48	7	47	12				
49	16	48	12				
50	15	49	12				
51	15	50	13				
				<div> <div>1000</div> <div>61,51</div> <div>448,33</div> <div>21,1</div> <div>76</div> </div>			
				<div> <div>$\binom{90}{2} = 4005$</div> <div>60,67</div> <div>444,40</div> <div>21,0</div> <div>79,01</div> </div>			
				<div> <div>1000</div> <div></div> <div></div> <div></div> <div></div> </div>			

TAB. IO. — *Frequenze dell'estremo superiore nell'insieme dei campioni, costituiti dai primi tre estratti di 1000 cinquine del lotto, e nel corrispondente universo in blocco.*

ESTREMO SUPERIORE	FRE- QUENZE NELL'IN- SIEME DEI 1000 CAM- PIONI	FREQUENZE NELL'UNIVERSO DEI CAMPIONI		ESTREMO SUPERIORE	FRE- QUENZE NELL'IN- SIEME DEI 1000 CAM- PIONI	FREQUENZE NELL'UNIVERSO DEI CAMPIONI	
		assolute	relative a 1000 cam- pioni			assolute	relative a 1000 cam- pioni
1	—	0		52	13	1275	11
2	—	0		53	14	1326	11
3	—	1		54	4	1378	12
4	—	3		55	11	1431	12
5	—	6		56	12	1485	13
6	—	10		57	10	1540	13
7	—	15		58	11	1596	14
8	—	21		59	12	1653	14
9	—	28	5	60	13	1711	15
10	—	36		61	9	1770	15
11	—	45		62	9	1830	16
12	—	55		63	15	1891	16
13	1	66		64	16	1953	17
14	—	78		65	17	2016	17
15	1	91		66	17	2080	18
16	—	105		67	19	2145	18
17	3	120		68	21	2211	19
18	3	136	5	69	15	2278	19
19	1	153		70	29	2346	20
20	3	171		71	24	2415	21
21	1	190		72	21	2485	21
22	2	210	5	73	26	2556	22
23	3	231		74	18	2628	22
24	1	253		75	23	2701	23
25	—	276	5	76	21	2775	24
26	5	300		77	25	2850	24
27	4	325	5	78	22	2926	25
28	3	351		79	20	3003	25
29	2	378	6	80	37	3081	26
30	4	406		81	31	3160	27
31	3	435	7	82	31	3240	28
32	7	465		83	24	3321	28
33	2	496	8	84	29	3403	29
34	3	528	5	85	32	3486	30
35	5	561	5	86	33	3570	30
36	9	595	5	87	32	3655	31
37	8	630	5	88	26	3741	32
38	4	666	6	89	38	3828	33
39	10	703	6	90	46	3916	33
40	7	741	6				
41	6	780	7				
42	5	820	7	Totale	1000	$\binom{90}{3} = 117.480$	1000
43	5	861	7	Media arit.	68,92	68,27	
44	11	903	8	Varianza	306,91	296,67	
45	4	946	8	Scarto quadratico medio	17,5	17,2	
46	5	990	8				
47	3	1035	8				
48	9	1081	9				
49	10	1128	10				
50	9	1176	10				
51	12	1225	10				
				Gradi di libertà: 64			
				$\chi^2 = 68,59$			

TAB. II. — *Frequenze dell'estremo superiore nell'insieme dei campioni, costituiti dai primi quattro estratti di 1000 cinquine del lotto, e nel corrispondente universo in blocco.*

ESTREMO SUPERIORE	FRE- QUENZE NELL'IN- SIEME DEI 1000 CAM- PIONI	FREQUENZE NELL'UNIVERSO DEI CAMPIONI		ESTREMO SUPERIORE	FRE- QUENZE NELL'IN- SIEME DEI 1000 CAM- PIONI	FREQUENZE NELL'UNIVERSO DEI CAMPIONI	
		assolute	relative a 1000 cam- pioni			assolute	relative a 1000 cam- pioni
1	—	0		52	12	20825	8
2	—	0		53	6	22100	9
3	—	0		54	4	23426	9
4	—	1		55	11	24804	10
5	—	4		56	9	26235	10
6	—	10		57	9	27720	11
7	—	20		58	12	29260	12
8	—	35		59	10	30856	12
9	—	56		60	8	32509	13
10	—	84		61	12	34220	13
11	—	120		62	8	35990	14
12	—	165		63	17	37820	15
13	—	220	5	64	13	39711	16
14	—	286		65	19	41664	16
15	—	364		66	15	43680	17
16	—	455		67	14	45760	18
17	—	560		68	24	47905	19
18	1	680		69	15	50116	20
19	—	816		70	30	52394	21
20	—	969		71	30	54740	22
21	—	1140		72	21	57155	22
22	—	1330		73	29	59640	23
23	—	1540		74	22	62196	24
24	1	1771		75	20	64824	25
25	—	2024		76	21	67525	26
26	2	2300		77	27	70300	28
27	2	2600		78	27	73150	29
28	1	2925	6	79	25	76076	30
29	—	3276		80	36	79079	31
30	1	3654		81	34	82160	32
31	1	4060		82	39	85320	33
32	5	4495	5	83	26	88560	35
33	2	4960		84	35	91881	36
34	1	5456		85	36	95284	36
35	2	5984	7	86	49	98770	38
36	4	6545		87	41	102340	40
37	4	7140	6	88	34	105995	41
38	3	7770	6	89	48	109736	43
39	4	8436	7	90	63	113564	44
40	4	9139					
41	3	9880	8	Totale	1000	$\binom{90}{4} = 2.555.190$	1000
42	2	10660		Media arit.	73,74	72,80	
43	4	11480	5	Varianza	205,31	208,98	
44	9	12341	5	Scarto quadratico medio	14,3	14,4	
45	2	13244	5				
46	6	14190	6				
47	2	15180	6				
48	5	16215	6				
49	6	17296	7				
50	7	18424	7				
51	5	19600	8				

Gradi di libertà = 54

$$\chi^2 = 60,29$$

TAB. 12. — *Frequenze dell'estremo superiore nell'insieme dei campioni, costituito da 1000 cinquine del lotto, e nel corrispondente universo in blocco.*

ESTREMO SUPERIORE	FRE- QUENZE NELL'IN- SIEME DEI 1000 CAM- PIONI	FREQUENZE NELL'UNIVERSO DEI CAMPIONI		ESTREMO SUPERIORE	FRE- QUENZE NELL'IN- SIEME DEI 1000 CAM- PIONI	FREQUENZE NELL'UNIVERSO DEI CAMPIONI	
		assolute	relative a 1000 cam- pioni			assolute	relative a 1000 cam- pioni
1	—	0		52	3	249900	6
2	—	0		53	5	270725	6
3	—	0		54	2	292825	7
4	—	0		55	11	316251	7
5	—	1		56	6	341055	8
6	—	5		57	7	367290	8
7	—	15		58	11	395010	9
8	—	35		59	8	424270	10
9	—	70		60	7	455126	10
10	—	126		61	15	487635	11
11	—	210		62	8	521855	12
12	—	330		63	10	557845	13
13	—	495		64	11	595665	14
14	—	715		65	20	635376	14
15	—	1001		66	13	677040	15
16	—	1365	5	67	13	720720	16
17	—	1820		68	18	766480	18
18	—	2380		69	12	814385	19
19	—	3060		70	30	864501	20
20	—	3876		71	28	916895	21
21	—	4845		72	15	971635	22
22	—	5985		73	24	1028790	23
23	—	7315		74	23	1088430	25
24	—	8855		75	32	1150626	26
25	—	10626		76	20	1215450	28
26	1	12650		77	27	1282975	29
27	—	14950		78	25	1353275	31
28	—	17550		79	28	1426425	32
29	1	20475		80	36	1502501	34
30	—	23751		81	38	1581580	36
31	—	27405		82	38	1663740	38
32	1	31465		83	30	1749060	40
33	—	35960		84	39	1837620	42
34	—	40920		85	44	1929501	44
35	2	46376	5	86	65	2024785	46
36	—	52360		87	47	2123555	48
37	3	58905		88	49	2225895	51
38	—	66045		89	59	2331890	53
39	—	73815	5	90	76	2441626	55
40	3	82251					
41	3	91390					
42	1	101270	7	Totale	1000	$\binom{90}{5} = 43.949.268$	1000
43	1	111930		Media arit.	76,73	75,83	
44	7	123410		Varianza	153,85	144,39	
45	4	135751	6				
46	3	148995					
47	2	163185	7	Scarto quadratico medio	12,4	12,0	
48	3	178365					
49	3	194580	8				
50	4	211876					
51	5	230300	10				

Gradi di libertà: 46
 $\chi^2 = 62,27$

quantitativamente il fenomeno sono stati calcolati i vari χ^2 , ottenendo i risultati :

- I) per i campioni di 2 unità : 76 gradi di libertà, $\chi^2 = 79,01$
- II) per i campioni di 3 unità : 64 gradi di libertà, $\chi^2 = 68,59$
- III) per i campioni di 4 unità : 54 gradi di libertà, $\chi^2 = 60,29$
- IV) per i campioni di 5 unità : 46 gradi di libertà, $\chi^2 = 62,27$.

Dalle tavole del χ^2 si ha che la probabilità di avere una distribuzione che si discosti da quella teoricamente determinata come quella trovata o ancora di più è certamente superiore a 0,100 nel primo caso, a 0,342 nel secondo, a 0,267 nel terzo e nell'ultimo caso a 0,057. Fissando il livello di significatività al 5 %, si può concludere che le differenze riscontrate non sono significative e quindi possono essere di natura puramente accidentale.

Nelle tabelle 9, 10, 11 e 12 sono state riportate anche le medie aritmetiche degli estremi superiori dei mille campioni e la rispettiva media aritmetica nell'universo, data quest'ultima dalla formula :

$$\bar{x}''^{(s)} = 90 - \frac{90 - v}{v + 1}.$$

L'accostamento fra queste medie è molto buono, come anche fra le varianze, di cui quella dell'universo dei campioni è data, per la [76], della formula

$$\sigma_s^2 = (90 - v) \left(\frac{181}{v + 1} - \frac{90 - v}{(v + 1)^2} - \frac{182}{v + 2} \right).$$

Dai valori riportati nelle tabelle si osserva che, aumentando il numero delle unità del campione, mentre la media aritmetica cresce e si avvicina allo estremo superiore della popolazione, lo scarto quadratico medio e la varianza diminuiscono sensibilmente.

26. — Per la distribuzione del campo di variazione, la [54] fornisce, nell'ipotesi che la popolazione sia anche equispaziata,

$$\varphi''(q, d) = \frac{(m - q + 1) \left[\binom{(q + 1) n'}{v} - 2 \binom{q n'}{v} + \binom{(q + 1) n'}{v} \right]}{\binom{(m + 1) n'}{v}}$$

che diviene, nel caso $n' = 1$,

$$\varphi''(q, d) = \frac{(N - q) \left[\binom{q+1}{v} - 2 \binom{q}{v} + \binom{q-1}{v} \right]}{\binom{N}{v}}$$

e dopo facili calcoli

$$\varphi''(q, d) = \frac{(N - q) \binom{q-1}{v-2}}{\binom{N}{v}},$$

formula, che non è altro che quella trovata dal Salvemini e riportata nel lavoro citato alla nota 2. Lo stesso autore ha determinato per la funzione di ripartizione l'espressione

$$\Phi''(q, d) = \frac{\binom{q+1}{v}}{\binom{N}{v}} \left[\frac{vN}{q+1} - (v-1) \right],$$

che può ottenersi come caso particolare (quando $n' = 1$) dalla seguente formula, ottenuta dalla [56],

$$\Phi''(q, d) = \frac{(m - q + 1) \binom{(q+1)n'}{v} - (m - q) \binom{qn'}{v}}{\binom{(m+1)n'}{v}},$$

che fornisce la funzione di ripartizione nel caso della popolazione equispaziata.

Per la media aritmetica del campo di variazione la [62] dà, ricordando l'identità

$$\binom{v}{v} + \binom{v+1}{v} + \dots + \binom{q}{v} = \binom{q+1}{v+1},$$

$$\overline{w''} - d \left[m - 2 \frac{\binom{m n' + 1}{v+1}}{\binom{(m+1)n'}{v}} \right]$$

che, nel caso $n' = 1$, si riduce alla

$$\overline{w''} = d \left[m - 2 \frac{N - v}{v + 1} \right] = d \frac{(v - 1)(N + 1)}{v + 1}.$$

Anche quest'ultima formula è stata trovata dal Salvemini, al cui lavoro, citato alla nota 2, si rimanda anche per le applicazioni numeriche in esso contenute.

BIBLIOGRAFIA

- GUMBEL E. J.: *Statistical Theory of Extreme Values and Some Practical Applications*. Nat. Bur. of Standards. Appl. Math. Series. 33. 1954.
- BURR. I.: *Calculation of exact sampling distribution of ranges from a discrete population*. Ann. Math. Stat. Vol. XXVI n. 3. 1955.
- CASTELLANO V.: *Cours de Sondage: théorie et méthodes*. Organisation des Nations Unies pour l'alimentation et l'agriculture. Roma. 1955.
- SALVEMINI T.: *Sui campioni di una massa discreta equidistribuita: il campo di variazione*. Statistica. Anno XVIII, n. 1. Gennaio-marzo 1958.
- GUMBEL, E.J.: *Statistics of Extremes*. Columbia University Press. New York. 1958.
- CASTELLANO V.: *Sull'universo dei campioni come fondamento della teoria dei campioni*. Bull. de l'Institut Int. de Statistique., T. XXXVII. 1960.

Si rimanda inoltre alla vasta bibliografia del lavoro del SALVEMINI per le pubblicazioni sul campo di variazione, e a quella del volume del GUMBEL (1958) per le pubblicazioni riguardanti gli estremi.

ALFREDO RIZZI

Sulle probabilità degli errori di compenso

SOMMARIO. — Determinati controlli sono necessari ma non sufficienti a garantire l'esattezza delle elaborazioni eseguite su dati statistici o contabili. Dopo aver trattato il problema in generale si studia la variabile casuale « somma di due numeri di k cifre » e si determina la probabilità degli errori di compenso nel caso che si controlli l'esattezza di n numeri di k cifre mediante confronto della loro somma con la somma esatta già nota e nel caso che si controlli l'esatta compilazione di una tabella a doppia entrata verificando che la somma degli elementi di ogni riga e di ogni colonna sia uguale alle frequenze marginali.

PREMESSA

Capita spesso che avendo dei dati statistici o contabili da elaborare si possa parzialmente controllare l'esattezza formale delle operazioni eseguite, mediante dei controlli.

Può darsi (e questo è, forse, il caso più comune) che questi controlli siano necessari ma non sufficienti a garantire l'esattezza delle elaborazioni eseguite come avviene, ad esempio, nella « prova del nove » della moltiplicazione e della divisione. Naturalmente per ogni « controllo » è molto importante conoscere qual'è la probabilità di accettare come esatto un risultato errato.

In questa nota verranno detti « errori di compenso » quelli non rilevati da un determinato controllo.

1. — Considerata una variabile discreta x che assume le determinazioni x_1, x_2, \dots, x_n , sia $f(x_1, x_2, \dots, x_n)$ una funzione reale di x_1, x_2, \dots, x_n di cui si conosce una sua determinazione a . Ad esempio, se si conosce la frequenza totale a di una seriazione stati-

stica, posto $f(x_1, x_2, \dots, x_n) = x_1 + x_2 + \dots + x_n$ deve essere $x_1 + x_2 + \dots + x_n = a$ se le x_i sono state esattamente rilevate. Così è noto che la media della variabile casuale descritta dalla media di tutti i possibili campioni, coincide con la media della massa dei dati da cui il campione è stato estratto, cioè

$$M(M_1) = \bar{x}$$

ove con M_1 si è indicata la variabile casuale (v.c.) media del campione e con \bar{x} la media dell'intera massa dei dati; se la relazione precedente non è verificata, diciamo che vi è un « errore » il quale può essere dovuto a molteplici cause (errori di trascrizione, errori di rilevazione, errori di calcolo, etc.).

In genere se per un qualsiasi motivo x_i diviene x_i^* sarà: $f(x_1, x_2, \dots, x_i^*, \dots, x_n) \neq a$ e diremo che l'errore che ha trasformato x_i in x_i^* è stato rilevato dal controllo $f(x_1, x_2, \dots, x_n) = a$.

2. — Se si indica con P_h la probabilità che vi siano h x_i errate tra le n determinazioni considerate (con $0 \leq h \leq n$) e si indica π_h la probabilità che essendovi h errori essi non siano rilevati dal controllo $f(x_1, x_2, \dots, x_n) = a$ (1) la probabilità che vi siano h errori non rilevati dal controllo considerato sarà:

$$\omega_h = P_h \cdot \pi_h \quad (1)$$

P_h dipende dalla probabilità p che la modalità x_i sia errata e, siccome rappresenta la probabilità che h tra le n modalità x_i siano errate, sarà:

$$P_h = \binom{n}{h} p^h \cdot q^{n-h} \quad (2)$$

(1) Potrà darsi che l'uguaglianza $f(x_1, x_2, \dots, x_n) = a$ si possa scrivere $f(x_1, x_2, \dots, x_n) = a + \varepsilon$ in quanto le x_1, x_2, \dots, x_n possono essere affette dagli errori $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$. È noto dalla teoria degli errori come si determina ε in funzione di $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ essendo noto il tipo di legame funzionale che lega x_1, x_2, \dots, x_n .

ove p è un valore che si ammette di conoscere per ogni particolare fenomeno; si suppone, cioè, che p sia una probabilità statistica. Supposto p « piccolo » (ad es., $\leq \frac{1}{100}$) la precedente si può scrivere

$$P_h = \frac{(np)^h e^{-np}}{h!} \quad (3)$$

Quindi

$$\omega_h = \frac{(np)^h e^{-np}}{h!} \pi_h \quad (4)$$

ove π_h , — che rappresenta, come si è detto, la probabilità che, essendovi h errori, essi si compensino — dipende da $f(x_1, x_2, \dots, x_n)$.

Lo schema precedente può trovare le più svariate applicazioni: ad esempio per controllare gli addendi di una somma mediante confronto del risultato con uno già noto, per controllare la esatta compilazione di una tabella a doppia entrata mediante la verifica che siano rispettati i vincoli orizzontali e verticali, per i controlli di validità in un elaboratore elettronico (2) etc. Ad esempio, nella pratica commerciale si usa controllare l'esattezza delle registrazioni contabili verificando che la somma di tutti i pagamenti positivi e negativi sia uguale allo sbilancio di cassa. In questo caso è utile conoscere qual'è la probabilità che due partite siano registrate errate, in maniera che gli errori siano tali che le due partite abbiano la somma esatta.

Nel seguito di questa nota vengono considerati due esempi

(2) Ammesso ad esempio che una qualsiasi cifra alfabetica o numerica sia sempre rappresentata da un numero pari di bit magnetici si vuole determinare la probabilità che perdendo o acquistando più « bit » questi sfuggano ad un controllo di parità. Analogamente se un carattere è rappresentato da un numero dispari di bit. Vedi, ad esempio: M. SCHATZAFF, W.B. HARDING: *A Mathematical model for determining the probabilities of undetected errors in magnetic tape system* « I.B.M. Journal of research and development » Aprile 1956.

che si possono assumere come casi particolari dello schema precedente.

3. — Siano n numeri reali non negativi di k cifre, di cui è nota la somma a . Indicata con p la probabilità che un numero di k cifre sia errato, la probabilità di due errori negli n numeri di k cifra sarà :

$$P_2 = \binom{n}{2} p^2 q^{n-2} \simeq \frac{(np)^2 e^{-np}}{2}$$

Si vuole determinare la probabilità che i due errori si compensino cioè la probabilità che sia $x_i + x_j = x_i^* + x_j^*$ ove con x_i^* e con x_j^* si sono indicati i valori « errati » corrispondenti a x_i ed x_j .

4. — Per lo scopo detto è opportuno premettere lo studio della v.c. S « somma di due numeri di k cifre » che si ottiene sommando a due a due tutti i possibili numeri di k cifre. La v.c. S assume le modalità

$$0, 1, 2, \dots, 10^k - 1, 10^k, 10^k + 1, \dots, 2(10^k - 1)$$

con probabilità rispettive

$$\frac{1}{10^{2k}}, \frac{2}{10^{2k}}, \frac{3}{10^{2k}}, \dots, \frac{10^k}{10^{2k}}, \frac{10^k - 1}{10^{2k}}, \dots, \frac{1}{10^{2k}}$$

in quanto i possibili numeri di k cifre sono 10^k , le coppie di possibili numeri sono 10^{2k} e si vede facilmente che le coppie di numeri che hanno la somma s sono

$$s + 1 \text{ se } 0 \leq s \leq 10^k - 1$$

$$2(10^k - 1) - s + 1 \text{ se } 10^k \leq s \leq 2(10^k - 1)$$

La v.c. S assume, quindi, la modalità s con probabilità

$$p(s) = \begin{cases} \frac{s+1}{10^{2k}} & \text{se } 0 \leq s \leq 10^k - 1 \\ \frac{2(10^k - 1) - s + 1}{10^{2k}} & \text{se } 10^k \leq s \leq 2(10^k - 1) \end{cases}$$

Ad esempio, per $K = 1$ la v.c. « S » assume le modalità

00, 01, 02, 03, ... 08, 09, 10, ... 17, 18

con probabilità rispettive

$$\frac{1}{10^2}, \frac{2}{10^2}, \frac{3}{10^2}, \frac{4}{10^2}, \dots, \frac{9}{10^2}, \frac{10}{10^2}, \frac{9}{10^2}, \frac{8}{10^2}, \dots$$

$$\dots, \frac{2}{10^2}, \frac{1}{10^2}$$

La v.c. « S » è quindi una v.c. simmetrica.

Ovviamente

$$\begin{aligned} \sum_{i=1}^{2(10^k-1)} p_i &= \frac{\sum_{s=1}^{10^k} s + \sum_{s=1}^{10^k-1} (10^k - 1 - s + 1)}{10^{2k}} = \\ &= \frac{\sum_{s=1}^{10^k-1} s + 10^k + \sum_{s=1}^{10^k-1} 10^k - \sum_{s=1}^{10^k-1} s}{10^{2k}} = \frac{10^k + 10^k(10^k - 1)}{10^{2k}} = 1 \end{aligned}$$

La moda della v.c. S è $10^k - 1$ e la probabilità della moda è

$$\frac{10^k}{10^{2k}} = \frac{1}{10^k}$$

La media aritmetica della v.c. « S » è pari a :

$$M(s) = \frac{\sum_{i=0}^{10^k-1} i(i+i) + \sum_{i=0}^{10^k} (2(10^k - 1) - i(i+i))}{10^{2k}}$$

con semplici passaggi si arriva a

$$M(s) = 10^k - 1$$

come si poteva ottenere anche tenendo conto che la v.c. S è la somma di due v.c. di media $\frac{10^k - 1}{2}$.

La media coincide quindi con la moda. Essendo inoltre $10^k - 1$ il valore mediano viene confermato che media, mediana e moda coincidono.

Per la varianza si ha:

$$\begin{aligned} \sigma_x^2 &= \frac{1}{10^{2k}} \left\{ \sum_{i=0}^{10^k-1} [i - (10^k - 1)^2 (i + i)] + \sum_{i=0}^{10^k-2} [2(10^k - 1) - i - \right. \\ &\quad \left. - (10^k - 1)^2 (i + 1)] \right\} \\ &= \frac{2}{10^{2k}} \sum_{i=0}^{10^k-2} [i - (10^k - 1) (i + 1)]. \end{aligned}$$

Sviluppando la sommatoria, dopo lunghi calcoli si ha:

$$\sigma_x^2 = \frac{1}{6} (10^{2k} - 1).$$

L'espressione precedente si poteva anche ottenere tenendo conto che la v.c. « S » è la somma di due v.c. indipendenti di varianza $\frac{10^{2k} - 1}{12}$ (3). Così ad esempio per $K = 1$ la media della v.c. « S » è 9, la varianza 16,5. Per $K = 2$ la media è 99 e la varianza 1666,5. Come curiosità si può dire che la media della v.c. « S » somma di due numeri di K cifre è formata da

(3) Vedi: C. GINI: *Variabilità e concentrazione*, 2ª edizione con note di T. SALVEMINI e E. PIZZETTI, Veschi, Roma, 1955.

un numero di tanti 9 per quanto è il valore di K , mentre la varianza è formata da un numero la cui prima cifra è 1 seguito da tanti sei per quanto è il doppio del valore di K meno 1 e da un unico decimale uguale a 5.

5. — Da quanto detto sulla v.c. S si desume che :

$$\frac{1}{10^{2k}} \leq \pi_2 \leq \frac{1}{10^k} \quad (5)$$

cioè la probabilità che, essendovi due errori, essi si compensino è minore o uguale ad $\frac{1}{10^k}$ (che è la probabilità della media, della moda e del valore mediano).

Sarà quindi :

$$\frac{(n p)^2}{2} e^{-np} \frac{1}{10^{2k}} \leq \omega_h \leq \frac{(n p)^2}{2} e^{-np} \frac{1}{10^k} \quad (6)$$

p si può chiamare « probabilità di errore dell'operatore » e dipende naturalmente anche da k , cioè dal numero delle cifre che compongono i singoli numeri. Per k non molto grande (≤ 10) generalmente si ammette

$$\frac{1}{100} \leq p \leq \frac{1}{1000}$$

La tabella seguente (4) fornisce il valore di

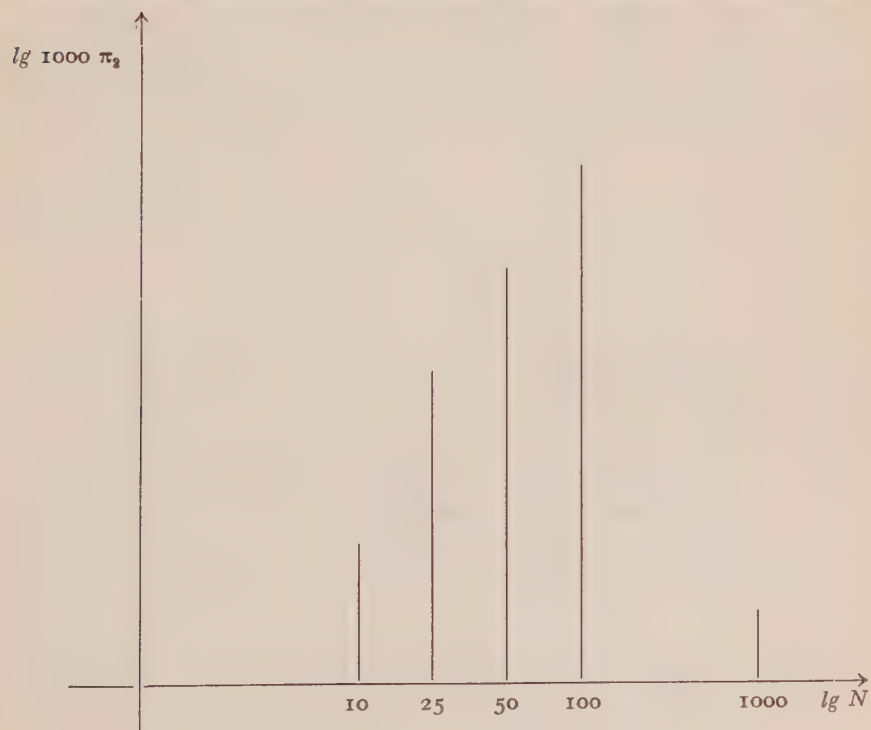
$$\pi_2 = \frac{(n p)^2}{2} e^{-np}$$

(4) Vedi: S. VIANELLI, *Metodologia statistica delle scienze agrarie*. Edizioni agricole, Bologna, Vol. 2°.

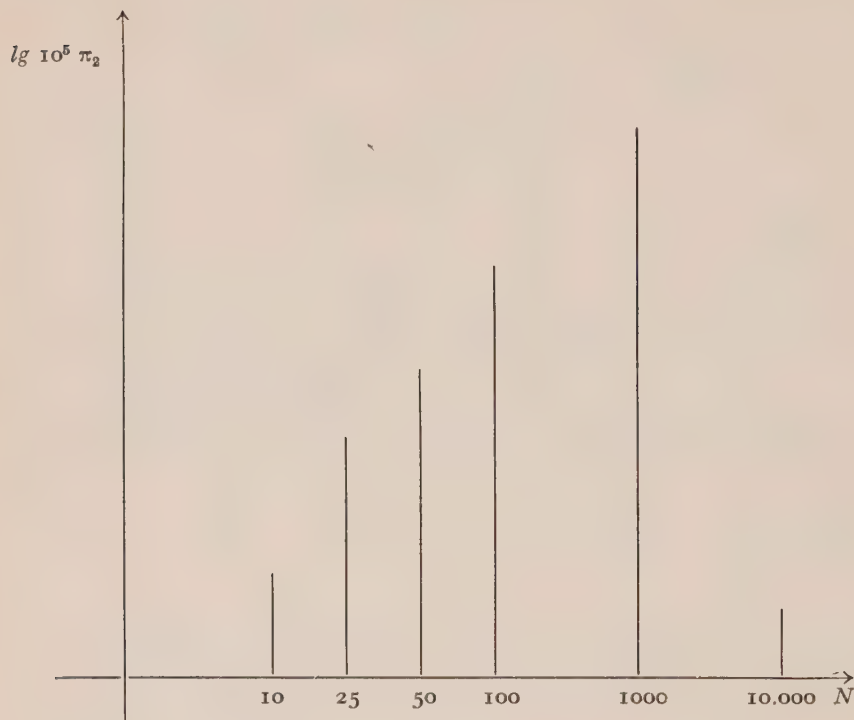
per alcuni valori di n e di p .

n	p	$\pi_2 = \frac{(n p)^2 e^{-np}}{2}$	n	p	$\pi_2 = \frac{(n p)^2 e^{-np}}{2}$
10	$\frac{1}{100}$	0,0045	10	$\frac{1}{1000}$	0,000049
25	»	0,024	25	»	0,00030
50	»	0,075	50	»	0,0011
100	»	0,183	100	»	0,0045
1000	»	0,0023	1000	»	0,1839
			10000	»	0,0023

Nel grafico seguente sono riportati i valori di π_2 relativi a $p = \frac{1}{100}$ in scala doppiamente logaritmica: sull'asse delle ascisse è riportato $\lg n$ e su quello delle ordinate $\lg 1000 \pi_2$.



Per $p = \frac{1}{1000}$ sull'asse delle ordinate è stato riportato $\lg 10^5 \pi_2$



Per ottenere i valori di ω_2 in corrispondenza di ogni k basterà moltiplicare i valori di π_2 riportati in tabella per 10^{-k} .
In genere sarà

$$\omega_h = P_h \cdot \pi_h \simeq \frac{(n p)^h}{h!} e^{-np} \pi_h \quad (7)$$

e

$$\frac{1}{10^{hk}} \leq \pi_h \leq \frac{1}{10^{(h-1)k}} \quad (8)$$

cioè la probabilità che essendovi h errori essi si compensino è :

$$\frac{(n p)^h}{h!} e^{-np} \frac{1}{10^{hk}} \leq \omega_h \leq \frac{(n p)^h}{h!} e^{-np} \frac{1}{10^{(h-1)k}} \quad (9)$$

APPLICAZIONE

In un centro meccanografico si perforano degli importi che si possono considerare di tre cifre significative su delle schede ad ottanta colonne. Gli altri elementi perforati, quali il numero di cliente, data ecc. non interessano ai fini di questa applicazione. Le perforazioni di questi importi si controllano mediante macchine meccanografiche, giacchè si conosce il totale dei primi quaranta addendi, dei secondi quaranta addendi ecc. che rappresentano gli importi che insieme ad altri dati sono contenuti in una pagina. Una rilevazione statistica ha messo in luce che la percentuale degli importi mal perforati è del 2,5 circa. Per cui la probabilità che su quaranta importi vi siano due errori tali che essi non siano rilevati confrontando la loro somma con quella nota è per la (8) :

$$\frac{(40 \cdot 0,025)^2}{2} e^{-40 \cdot 0,25} \frac{1}{10^3} \cong 183 \cdot 10^{-6}$$

Gli errori di compenso sono rilevati da un ulteriore controllo giacchè ogni elemento è assegnato ad un « cliente » e si conosce il « saldo » per cliente. Su circa centomila somme corrispondenti a lavoro del « centro » di un anno è stato riscontrato che ventitre errori non sono stati rilevati dal primo controllo mentre il numero medio di errori secondo la formula precedente deve essere

$$N P \cong 10^{-6} \cdot 183 \cdot 10^5 = 18,3$$

Lo scarto quindi è circa 4,7 e ammesso che la distribuzione ω_h sia normale sarà: $\sigma_x = \sqrt{NPq} \cong \sqrt{NP} = 4,2$. Nel caso considerato vi è conformità tra ipotesi e risultati sperimentali.

8 Sia data una tabella a doppia entrata

			N_{10}
	N_{ij}		N_{i0}
			N_{r0}
N_{01}	N_{0j}	N_{s0}	N

Come controllo della esatta compilazione della tabella precedente si deve verificare che la somma degli elementi della riga — i -ma deve essere N_{i0} e la somma degli elementi della colonna j -ma deve essere N_{0j} . Può darsi, però, che vi siano uno e più errori di compenso non rilevabili dal controllo che la somma degli elementi della riga i -ma sia N_{i0} e la somma degli elementi della colonna j -ma sia N_{0j} . Sia ad esempio la tabella

3	(2)	(1)	6
4	(9)	(5)	18
1	6	3	10
8	17	9	34

La tabella

3	(1)	(2)	6
4	(10)	(4)	18
1	6	3	10
8	17	9	34

verifica gli stessi vincoli della precedente.

Se vi sono un numero uguale a 2 o dispari di errori essi saranno rilevati dai controlli detti precedentemente, mentre possono non essere rilevati se il numero degli errori è quattro e gli elementi errati hanno una disposizione del tipo

	X		X	
	X		X	

(a 1)

oppure gli elementi errati sono sei ed hanno una disposizione del tipo

	X		X		
	X				X
			X		X

(a 2)

e così via.

Per determinare la probabilità che, essendovi quattro errori su due colonne determinate, essi non siano rilevati dal controllo orizzontale e verticale (probabilità che sarà indicata con ω'_4) bisognerà moltiplicare la probabilità di due errori di compenso su una determinata colonna per la probabilità di due errori di compenso su una colonna diversa dalla precedente ma sulle stesse righe.

Se si ammette che il numero appartenente alla cella ($i-j$) è di k cifre ed r è il numero di elementi appartenenti alla colonna prefissata (cioè il numero delle righe) sarà:

$$\omega'_4 = \omega_2 p^2 \frac{1}{10^{2k}} = \frac{(rp)^2}{2} e^{rp} \frac{1}{10^{4k}} p^2 \quad (10)$$

La probabilità che essendovi quattro errori nella tabella essi non siano rilevati dal controllo orizzontale e verticale sarà:

$$\binom{s}{2} \binom{s}{2} (\omega'_4)^1 (1 - \omega'_4)^{\binom{s}{2} - 1} \quad (11)$$

ove s è il numero delle colonne della tabella.

La precedente si può scrivere

$$\binom{s}{2} \left[\frac{(rp)^2}{2} e^{rp} \frac{1}{10^{4k}} p^2 \right] \cdot \left[1 - \omega'_4 \frac{s(s-1)-2}{2} \right] \cong$$

$$\frac{s \frac{(rp)^2}{2} e^{rp} \frac{1}{10^{4k}} p^2}{2} e^{-\frac{s(rp)^2}{2}} e^{rp} \frac{1}{10^{4k}} p^2$$

Considerazioni analoghe valgono nel caso della configurazione (a2).

CONCLUSIONE

Oggi, con l'introduzione di sistemi automatici, in molti settori si fa largo uso di sistemi di controllo globale. Occorre quindi sapere qual'è la probabilità che un errore sfugga ai sistemi di controllo nei vari processi. Naturalmente maggiori « vincoli » comportano necessariamente maggiori spese e quindi l'introduzione di un ulteriore controllo va valutata in funzione del maggiore costo che comporta e del minor danno che ne deriva.

AMATO HERZEL

Sul « problema del commesso viaggiatore »

- 1°. Enunciazione del problema. I metodi tradizionali per la sua soluzione. —
2°. Nuovi procedimenti per l'eliminazione dei percorsi non ottimi.
L'applicazione della programmazione lineare. — 3°. Applicazioni ad
un esempio: il percorso più breve fra 16 capoluoghi italiani.

1) Col termine, alquanto pittoresco, di « problema del commesso viaggiatore » (*travelling salesman problem*) si designa uno di quei problemi di ottimo condizionato che negli ultimi tempi hanno attirato su di sé un'attenzione che, ancora pochi anni fa, sarebbe stato difficile prevedere, tanto da dare vita ad un nuovo ramo della matematica applicata, noto sotto il nome di « ricerca operativa ». Più precisamente, si tratta, nel problema in esame, di trovare il più breve dei percorsi chiusi che uniscono un certo numero, diciamo n , di punti, P_1, P_2, \dots, P_n di cui sono note le distanze reciproche.

I dati del problema sono perciò le quantità d_{ij} ($i, j = 1, 2, \dots, n$) che, almeno nella versione più semplice del problema, soddisfano alle note proprietà delle distanze :

$$\begin{aligned} d_{ij} &= 0 \quad \text{se, e solo se} & i &= j \\ d_{ij} &\leq d_{ik} + d_{jk} & (i, j, k, &= 1, 2, \dots, n) \end{aligned}$$

e quindi :

$$d_{ij} = d_{ji} \geq 0 \qquad (i, j = 1, 2, \dots, n)$$

Nel seguito supporremo che tali proprietà siano soddisfatte e supporremo altresì che nella seconda relazione valga per qualunque coppia i, j , almeno per un valore di k il segno di disuguaglianza ; altrimenti, infatti il problema sarebbe banale (come

avverrebbe, per esempio, se le distanze venissero misurate lungo un tratto di curva o se i punti fossero allineati).

Se introduciamo le $n(n-1)$ variabili x_{ij} ($i \neq j$) alle quali, in relazione ad un percorso dato, facciamo assumere il valore 1 oppure 0, a seconda che nel percorso, figuri, o no, il tratto P_i-P_j , oppure il tratto P_j-P_i , il nostro problema consiste nel determinare i valori delle variabili x_{ij} in modo tale da rendere minima la funzione :

$$D = \frac{1}{2} \sum_{ij} d_{ij} x_{ij} \quad (i \neq j) \quad (1)$$

$$\text{oppure } D = \sum_{ij} d_{ij} x_{ij} \quad (i < j) \quad (2)$$

subordinatamente alle condizioni :

$$\sum_j x_{ij} = 1, \quad (i = 1, 2, \dots, n; j \neq i) \quad (3)$$

rispettivamente :

$$\sum_{j=i+1}^n x_{ij} + \sum_{h=1}^{i-1} x_{hi} = 2 \quad (4)$$

e subordinatamente alla condizione, difficile ad esprimersi in simboli matematici, che il percorso non si scomponga in subpercorsi chiusi.

È appunto, quest'ultima condizione che rappresenta una delle difficoltà maggiori, per cui anche il tentativo di risolvere il problema coi metodi ormai tradizionali della programmazione lineare (Dantzig) ha avuto un successo solo parziale. Difatti, allo stato attuale non si conoscono regole generali per tradurre la condizione suddetta in un certo numero, non troppo elevato, di vincoli lineari, cosa questa necessaria per la pratica applicazione del metodo del semplice o degli altri metodi della programmazione lineare.

Un'altra difficoltà è costituita dalle soluzioni nelle quali le variabili x_{ij} assumono valori frazionari. Non di rado sono proprio quelle le soluzioni che danno alla funzione D il suo valore minimo. Per evitare tali soluzioni frazionarie, è talvolta necessario introdurre nuovi vincoli lineari, la cui formulazione spesso non è affatto immediata ed evidente e che, comunque, accrescono la

mole di lavoro occorrente per giungere alla individuazione del percorso ottimo. Con ciò non si vuol dire che la tecnica della programmazione lineare sia del tutto impotente di fronte al « problema del commesso viaggiatore ». Anzi, è stato con l'ausilio, se non proprio per mezzo esclusivo, di tali metodi che G. Dantzig e altri hanno risolto il problema per 42 città americane, il che rappresenta tuttora una specie di primato, in quanto non si ha notizia di altri problemi. con un numero uguale o maggiore di punti, che siano stati risolti.

Occorre tuttavia osservare che, nell'esempio citato, i metodi della programmazione lineare non risultano, di per sè, sufficienti per dimostrare che il percorso trovato sia quello ottimo. Invero, essi, in definitiva, riescono solo a dimostrare che numerosi tratti non possono figurare nella soluzione ottima. In tal modo, si riduce fortemente l'insieme dei percorsi fra i quali occorre cercare quello più breve.

Anche gli altri metodi di soluzione che attualmente si conoscono, si fondano sostanzialmente sull'eliminazione di gruppi possibilmente numerosi di percorsi, di cui sia possibile dimostrare che non contengono quello ottimo, senza che sia necessario calcolare la lunghezza dei percorsi da eliminare.

L'utilità, e anzi la necessità pratica di questi procedimenti poggia sul fatto che, se il numero dei punti non è piccolissimo, nemmeno le più moderne calcolatrici elettroniche potrebbero calcolare la lunghezza di tutti i percorsi possibili, in un tempo utile per il conseguimento di un qualunque scopo concreto, poichè com'è facile vedere, il numero complessivo dei percorsi chiusi possibili è uguale a $\frac{1}{2} (n - 1)!$. Anche per n non eccessivamente grande, si tratta di un numero elevatissimo, tale da rendere del tutto inapplicabile il procedimento suddetto, anche per chi può disporre delle velocità consentite dal calcolo elettronico. Di qui la importanza dei metodi che riescono a ridurre il numero dei percorsi fra cui ricercare il percorso ottimo.

A tale uopo ci si serve generalmente di alcuni fatti noti (pochi per la verità), di cui i principali sono i seguenti:

a) la linea del percorso ottimo non può intersecare se stessa. Tale proprietà, già di per se abbastanza evidente e, del resto, facilmente dimostrabile, viene utilizzata soprattutto nei tentativi

di trovare soluzioni per via grafica: partendo da un qualunque percorso, si giunge a un percorso migliore eliminando le intersezioni. Per questa via si riesce spesso a giungere a buoni percorsi, purtroppo non appare possibile dimostrare in tal modo che un determinato percorso sia quello ottimo.

Non risulta che tale proprietà sia stata finora utilizzata per una sistematica eliminazione di percorsi che sicuramente non sono ottimi;

b) la proprietà predetta è compresa nella seguente, che rappresenta un requisito più restrittivo: supposto che nel percorso ottimo in cui sia fissato un orientamento, figurino i tratti $P_i - P_j$ e $P_k - P_l$, si dovrà verificare la seguente disuguaglianza:

$$d_{ij} + d_{kl} \leq d_{ik} + d_{jl};$$

c) una proprietà simile è quella espressa dalla seguente relazione che vale per il percorso ottimo: supposto che vi figurino i tratti $P_k - P_i - P_j$ e $P_k - P_l$ si deve avere:

$$d_{ki} + d_{ij} - d_{kj} \leq d_{kl} + d_{le} - d_{kl}$$

Anche tale proprietà, sulla quale, come sulla precedente, avremo modo di tornare estesamente in seguito, viene utilizzata soprattutto per migliorare in modo rapido e semplice, un percorso.

Non risulta, invece, che essa sia stata sfruttata per una eliminazione sistematica di percorsi non ottimi. Di conseguenza, al pari della proprietà precedente, anche questa, allo stato attuale, appare più utile ai fini della ricerca di una soluzione buona, (che, incidentalmente, può anche essere quella ottima) che non ai fini dell'individuazione della soluzione ottima, oppure della dimostrazione che una soluzione determinata è quella ottima. A questo proposito si può osservare che proprio tale dimostrazione rappresenta, in realtà, l'aspetto più arduo del problema, in quanto, per ciò che riguarda la ricerca empirica di un buon percorso la semplice ispezione visiva di una pianta su cui sia riprodotta la dislocazione dei punti, integrata con metodi fondati sulle proprietà esposte, permette talvolta, soprattutto se il numero dei punti è poco elevato, di individuare con una certa rapidità dei percorsi

che, dal punto di vista pratico, possono essere considerati quasi equivalenti a quello ottimo ;

d) La seguente proprietà, invece, è più utile ai fini di una ricerca rigorosa della soluzione ottima, che non a quelli dell'individuazione di un percorso breve. Si tratta di una proprietà semplice, ma importante, che può essere così formulata : la soluzione ottima non muta se da tutti gli elementi di una linea della matrice delle distanze (riga o colonna) si sottrae una quantità costante, e non varia nemmeno se si compie tale operazione contemporaneamente per tutte le linee. In simboli, cioè, si ha che, se alle distanze $d_{i,j}$ si sostituiscono le quantità $\delta'_{i,j}$ così definite :

$$\delta'_{i,j} = d_{i,j} - u_i - u_j \quad (i, j = 1, 2, \dots, n, i \neq j) \quad (5)$$

dove le u sono quantità arbitrarie, il percorso ottimo rimane invariato, poichè, in virtù della (3) o della (4), con tali modificazioni viene semplicemente aggiunta o tolta una stessa quantità costante, il che, come è noto, non altera il problema del minimo. Con una opportuna scelta delle costanti u , tale proprietà può rivelarsi molto importante : talvolta può avvenire perfino che si annullino tutte le $\delta_{i,j}$ corrispondenti a tratti utilizzati in un determinato percorso (ai quali fanno, cioè, riscontro, variabili $x_{i,j}$ aventi il valore 1) mentre tutte le altre δ risultino positive o, almeno, non negative ; allora è chiaro che il percorso dato è quello ottimo, o, per lo meno, è uno degli ottimi. Questo caso, purtroppo, è eccezionale, e si verifica soltanto in problemi del tutto particolari ; ma è, invece, pienamente possibile, soprattutto se il numero dei punti non è troppo elevato, che la lunghezza complessiva di un determinato percorso, espressa in distanze trasformate $\delta_{i,j}$, risulti inferiore ai valori di alcune altre δ . Allora, supposto che nella matrice delle δ non figurano quantità negative, si possono senza altro scartare i tratti corrispondenti a tali δ e, quindi, tutti i percorsi in cui quei tratti figurano. Ciò è tanto più importante, in quanto, variando i valori delle u , si possono spesso eliminare altri tratti e quindi altri percorsi. Anche l'applicazione della programmazione lineare al nostro problema, applicazione di cui si è fatto cenno prima, si fonda in gran parte sul procedimento ora delineato.

Un semplice esempio può servire ad illustrare in qualche modo alcune delle cose dette finora. Siano dati sei punti, (che, per maggior semplicità, indicheremo con i numeri, come faremo anche in seguito), di cui diamo la rappresentazione grafica (Fig. 1) e la matrice di valori approssimativi delle distanze d_{ij} (Tab. n. 1).

TABELLA N. 1. — *Matrice delle distanze d_{ij}*

	1	2	3	4	5	6
1	—	2	2,6	2	1	2
2	2	—	1	2,6	2	3,7
3	2,6	1	—	2	2	3,7
4	2	2,6	2	—	1	2
5	1	2	2	1	—	1,9
6	2	3,7	3,7	2	1,9	—



Fig. 1

Si supponga ora di aver scelto, come tentativo iniziale il percorso : 1,3,5,4,2,6,(1) (Fig. 2), la cui lunghezza complessiva è pari a 13,9.

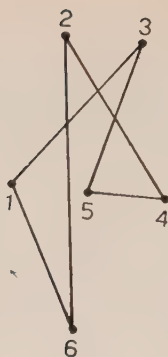


Fig. 2

In base alla proprietà *a*), si può passare per esempio, al percorso 1-5-3-4-2-6-(1), (Fig. 3) che contiene solo 2 intersezioni,

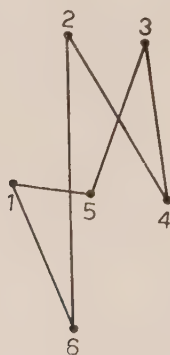


Fig. 3

anzichè 3 e la cui lunghezza è pari a 13,3. Alla stessa modificazione si può giungere in base al requisito *b*) ; infatti, si ha :

$$d_{1,3} + d_{4,5} = 2,6 + 1 = 3,6 > d_{1,5} + d_{3,4} = 1 + 2 = 3$$

Infine, anche in base alla proprietà *c*) si può passare allo stesso percorso, avendosi : $d_{1,3} + d_{3,5} - d_{1,5} = 2,6 + 2 - 1 = 3,6 >$
 $d_{3,5} + d_{3,4} - d_{4,5} = 2 + 2 - 1 = 3$

Facendo ricorso alla proprietà *a*) o *c*), si può passare quindi, per esempio, al percorso 1,2,5,3,4,6,(1), di lunghezza 12 (Fig. 4). Il requisito *a*)



Fig. 4

è ora soddisfatto, non così i requisiti *b*) e *c*), per soddisfare il primo dei quali si può passare per esempio al percorso 1,5,2,3,4,6,(1) avente lunghezza 10 (Fig.5). Le condizioni *a*) e *b*) sono ora soddisfatte non così la condizione *c*); infatti: $d_{1,5} + d_{2,5} - d_{1,2} = 1 + 2 - 2 = 1 > d_{4,5} + d_{5,6} - d_{4,6} = 1 + 1,9 - 2 = 0,9$.

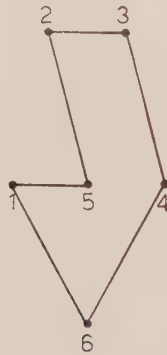


Fig. 5

Si giunge così ai percorsi 1,2,3,4,5,6,(1) e 1,2,3,4,6,5,(1) aventi entrambi lunghezza 9,9 (Fig. 6a) e (Fig. 6b).

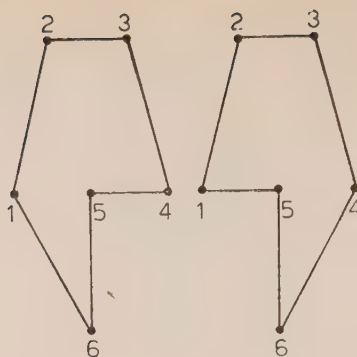


Fig. 6 a

Fig. 6 b

Sono finalmente soddisfatti tutti i requisiti sopra esposti, come non è difficile verificare. Appare pertanto lecito ritenere che i due percorsi ora trovati siano i più brevi. Tuttavia, non si può ancora affermarlo con certezza, in quanto le tre condizioni soddisfatte sono bensì necessarie, ma non sono assolutamente sufficienti. Quando il numero dei punti è più elevato, non è anzi difficile trovare parecchi percorsi non ottimi, per i quali le condizioni suddette siano soddisfatte.

Per tentare di dimostrare che i percorsi trovati siano effettivamente i migliori, si può ricorrere alla proprietà *d*). Ponendo,

TABELLA N. 2

	1	2	3	4	5	6	<i>u</i>
1	—	0	1,6	0	0	0	1
2	0	—	0	0,6	1	1,7	1
3	1,6	0	—	1	2	2,7	0
4	0	0,6	1	—	0	0	1
5	0	1	2	0	—	0,9	0
6	0	1,7	2,7	0	0,9	—	1
<i>u</i>	1	1	0	1	0	1	

per esempio, $u_1 = 1$, $u_2 = 1$, $u_3 = 0$, $u_4 = 1$, $u_5 = 0$, $u_6 = 1$ (torneremo più tardi sui metodi per calcolare i valori più convenienti delle quantità u), si ottiene, quale matrice delle distanze trasformate ($\delta_{ij} = d_{ij} - u_i - u_j$) la tab. n. 2.

La lunghezza dei percorsi 1,2,3,4,5,6, (1) e 1,2,3,4,6,5,(1) espressa in distanze trasformate, è pari a 1,9. Poichè la matrice di tali distanze non contiene elementi negativi, è chiaro che tutti i percorsi recanti i tratti 3-5 e 3-6, ai quali corrispondono distanze ridotte superiori a 1,9 (2 e 2,7 rispettivamente) risulteranno più lunghi dei percorsi indicati. Di conseguenza, si possono trascurare tutti i percorsi contenenti i tratti 3-5 e 3-6. Cambiando ora i valori di u_2 e di u_3 e precisamente ponendo: $u_2 = 0$, $u_3 = 1$, mentre rimangono invariati i valori delle altre u , si ottiene per le distanze trasformate δ_{ij} la n. 3.

TABELLA N. 3

	1	2	3	4	5	6	u
1	—	1	0,6	0	0	0	1
2	1	—	0	1,6	2	2,7	0
3	0,6	0	—	0	1	1,7	1
4	0	1,6	0	—	0	0	1
5	0	2	1	0	—	0,9	0
6	0	2,7	1,7	0	0,9	—	1
u	1	0	1	1	0	1	

La lunghezza dei percorsi, in base alla Tab. n. 3 è invariata, rispetto alla tabella precedente, essendo rimasta invariata la somma delle u_i . Ora però risulta che possono essere eliminati anche i tratti 2-5 e 2-6 ai quali corrispondono distanze trasformate superiori a 1,9. Modificando ancora i valori delle u_i in modo, sempre, da lasciarne invariata la somma, si giunge alla tabella n. 4.

TABELLA N. 4

	1	2	3	4	5	6	u
1	—	0	2,5	0,9	0	0	0,55
2	0	—	0	0,6	0,1	0,8	1,45
3	2,5	0	—	1,9	2	2,7	-0,45
4	0,9	0,6	1,9	—	0	0	0,55
5	0	0,1	22	0	—	0	0,45
6	0	0,8	2,7	0	0	—	1,45
u	0,55	1,45	-0,45	0,55	0,45	1,45	

In base alla tabella n. 4, diviene immediata anche la eliminazione del tratto 1-3. Poichè il punto 3 deve essere collegato con altri due punti, mentre risultano esclusi i tratti 1-3, 5-3, e 6-3, è evidente che ogni percorso ottimo, deve contenere i tratti 2-3-4. Con ciò viene ad essere eliminato automaticamente il tratto 2-4, poichè altrimenti si verrebbe a creare un sub-percorso chiuso. Dato che, in base alla tabella n. 3, i tratti 2-5 e 2-6 sono eliminati, il percorso ottimo deve contenere anche il tratto 1-2.

Riassumendo, si è così dimostrato che in ogni percorso ottimo devono figurare i tratti 1-2-3-4. Risulta quindi dimostrato che i due percorsi già individuati sono effettivamente gli unici percorsi ottimi.

Il lettore potrà osservare, sulla base di questo esempio, che il « problema del commesso viaggiatore » presenta difficoltà piuttosto rilevanti, anche quando il numero dei punti è così esiguo. È vero che il procedimento per l'indicazione del percorso ottimo avrebbe forse potuto essere abbreviato se si fosse partito con un percorso migliore, come sarebbe stato facile trovarne. È vero, anche, che la seconda parte del procedimento risolutivo, cioè la prova che i percorsi trovati sono ottimi, avrebbe potuto basarsi esclusivamente sulle tabelle n. 3 e n. 4 in quanto la tabella n. 2 non permette nessuna eliminazione che non sia giustificata anche in base alla tabella n. 4. D'altra parte, però, le difficoltà aumentano rapidamente col crescere del numero dei punti. Non appare

perciò fuori di luogo la ricerca di altri espedienti o metodi che possano agevolare la soluzione del problema. E ciò sarà fatto nei paragrafi seguenti, dove verrà anche risolto, a titolo di esempio un problema di medie dimensioni riguardante l'Italia.

Nel concludere il presente paragrafo, dedicato ad una rapida rassegna, dei principi più importanti, su cui si fondano attualmente i procedimenti di soluzione, rinviano il lettore, desideroso di approfondire l'argomento agli articoli citati nella bibliografia.

2) Come si è visto nel paragrafo precedente, vi è attualmente una distinzione abbastanza netta fra i metodi tendenti ad individuare un buon percorso, possibilmente anzi il migliore, e quelli destinati invece a sgombrare il campo da percorsi certamente non ottimi mediante l'eliminazione di certi tratti. Mostriamo ora come sia possibile sfruttare anche le proprietà *b)* e *c)* ai fini dell'eliminazione suddetta e, quindi, della dimostrazione rigorosa che un dato percorso sia ottimo.

Prendiamo anzitutto la proprietà *b)*. Come si è visto, essa si fonda su un fatto molto evidente che può essere così enunciato: dato un percorso $P_1 \dots P_{i_1} - P_{i_2} - P_{i_3} \dots - P_{i_{k-1}} - P_{i_k} - P_{i_{k+1}} \dots (P_1)$ se sussiste la disuguaglianza:

$$d_{i_1 i_2} + d_{i_k i_{k+1}} > d_{i_1 i_k} + d_{i_2 i_{k+1}} \quad (6)$$

si ottiene un percorso più breve operando una modificazione semplicissima, che si può chiamare « inversione »; si passa cioè al percorso $P_1 \dots P_{i_1} - P_{i_k} - P_{i_{k-1}} \dots P_{i_3} - P_{i_2} - P_{i_{k+1}} \dots (P_1)$. Consideriamo ora due righe (o colonne) della matrice delle distanze (d_{ij}) , per esempio le righe *k* esima o, più brevemente, riga *k* e *l* esima, o riga *l*. Calcoliamo le differenze fra gli elementi corrispondenti della riga *k* e della riga *l* cioè fra gli elementi di eguale colonna. Indichiamo col simbolo ${}_1P_{kl}$ il punto al quale corrisponde il massimo valore di dette differenze, cioè il massimo valore della espressione:

$${}_j\Delta_{k,l} = d_{kj} - d_{lj} \quad (7)$$

al variare di j . Si abbia, tanto per fissare le idee :

${}_1P_{kl} = P_h$, cioè :

$$\text{Max}_j {}_j\Delta_{k,l} = {}_h\Delta_{k,l}$$

In base a quanto esposto sopra, è allora evidente che il percorso ottimo può contenere il tratto $P_k - P_h$ soltanto se contiene anche il tratto $P_h - P_l$.

Diversamente si potrebbe sempre giungere ad un percorso più breve mediante un'inversione che sostituisca al tratto $P_h - P_k$ il tratto $P_h - P_l$: Infatti, per qualunque valore di $m \neq h, k$, il percorso $P_1 \dots P_h - P_k - \dots P_l - P_m \dots - (P_1)$ risulta più lungo del percorso

$$P_1 \dots P_h - P_l \dots P_k - P_m \dots (P_1)$$

Possiamo perciò dire che il tratto $P_k - P_h$ è sub-eliminato dal punto P_l . Così per esempio, se si fa il confronto delle righe 1 e 2 della tabella n. 1 il tratto 1-3 risulta sub-eliminato dal punto 2 e il tratto 2-6 risulta sub-eliminato dal punto 1.

Effettuando i confronti suddetti fra tutte le coppie di righe, molte sub-eliminate si trasformano di solito in eliminazioni complete. Ciò può avvenire in tre modi diversi :

a) se si ha per esempio: $P_h = {}_1P_{k,l} = {}_1P_{k,m}$ il tratto $P_h - P_k$ non può figurare nel percorso ottimo, che altrimenti dovrebbe contenere anche i tratti $P_h - P_l$ e $P_h - P_m$ il che è impossibile. Questo tipo di eliminazione ricorre frequentemente ; per tornare al nostro esempio il tratto 1-3 è eliminato in quanto sub-eliminato da 2 e da 4.

b) Se si ha : $P_h = {}_1P_{h,l}$ e $P_k = {}_1P_{k,l}$ il tratto $P_h - P_k$ non può, ovviamente, essere compreso nel percorso ottimo, altrimenti dovrebbero figurare contemporaneamente i tratti $P_l - P_k - P_h$ e $P_k - P_h - P_l$, il che è impossibile, se, come è naturalmente supposto, $n > 3$. Ogni punto dà luogo ad almeno una eliminazione

di questo tipo, e precisamente, fissato comunque l'indice l , viene eliminato il tratto $P_h - P_k$ per quei valori di h e k che rendono minima l'espressione :

$$d_{h,l} + d_{k,l} - d_{h,k} = d_{h,l} + {}_k\Delta_{l,h} = d_{k,l} + {}_k\Delta_{l,k} \quad (8)$$

espressione che in seguito indicheremo col simbolo (l, h, k) .

Così, nel nostro esempio, il punto 1 elimina il tratto 2-6, poichè $P_2 = {}_1P_{6,1}$ e $P_6 = {}_1P_{2,1}$, il punto 2 elimina il tratto 1-3, poichè $P_1 = {}_2P_{3,2}$ e $P_3 = {}_2P_{1,2}$, ecc. ecc.

c) Un altro caso, assolutamente ovvio, di trasformazione di una sub-eliminazione in eliminazione si ha quando il tratto che potrebbe rendere possibile quello sub-eliminato, risulta a sua volta eliminato. Così, nel nostro esempio, con un'opportuna convenzione di cui si farà cenno più sotto, può risultare : ${}_1P_{6,5} = P_3$ ossia il tratto 3-6 è sub-eliminato, potendo figurare nel percorso ottimo solo se vi figura anche il tratto $P_3 - P_5$. Quest'ultimo tratto, però, viene eliminato nei confronti della riga 5 con le righe 2 e 4, in quanto si ha : ${}_1P_{5,2} = {}_1P_{5,4} = P_3$. Di conseguenza, anche il tratto 3-6 è eliminato.

Dal fatto sopra illustrato, deriva inoltre che in situazioni di quel genere i confronti fra certe coppie di righe possono essere ripetuti e condurre successivamente all'eliminazione di diversi tratti. Ciò in quanto, per la possibilità di « inversione » sono irrilevanti i punti che nel percorso ottimo non possono essere uniti da tratti diretti con quelli corrispondenti alle righe fra cui si fanno i confronti. In simboli, se designano con I_l l'insieme dei punti che, in un determinato stadio del procedimento risolutivo, possono essere uniti da tratti diretti con P_l e se, per l qualunque, ${}_1P_{k,l}$ non è compreso in I_l , si può senz'altro eliminare il tratto ${}_1P_{k,l} - P_k$ e, in pari tempo, sub-eliminare il tratto $P_k - {}_2P_{k,l}$ (dove ${}_2P_{k,l}$ sta a designare il punto al quale corrisponde il secondo massimo di ${}_j\Delta_{k,l}$ al variare di j), purchè sia compreso in I_l . Se ciò non si verifica, si elimina anche il tratto $P_k - {}_2P_{k,l}$ e si passa a condiderare il punto ${}_3P_{k,l}$ e così via. Nel nostro esempio, vengono in tal modo eliminati, a determinate condizioni, mediante il confronto 6-5, integrato dai confronti 5-1, 5-2, 5-3, 5-4, successivamente i tratti 6-3 e 6-2 e, quindi, può venire sub-eliminato il tratto 6-1 oppure il tratto 6-4.

Si vede immediatamente che i confronti fra le righe, su cui si fondano le eliminazioni sopra descritte, possono essere condotti, indifferentemente, anche sulle tabelle delle distanze trasformate, δ_{ij} . Infatti, per queste eliminazioni hanno rilevanza soltanto le graduatorie fra le differenze ${}_j\Delta_{k,l}$ e tali graduatorie non vengono modificate quando si passa alle distanze trasformate, data la relazione :

$$\delta_{kj} - \delta_{lj} = (d_{kj} - u_k - u_j) - (d_{lj} - u_l - u_j) = {}_j\Delta_{k,l} + u_l - u_k.$$

Da tale relazione si desume che la differenza fra $\delta_{kj} - \delta_{lj}$ e $d_{kj} - d_{lj}$ non varia al variare di j . Di conseguenza la graduatoria delle quantità $\delta_{kj} - \delta_{lj}$ al variare di j , coincide con quella della quantità $d_{kj} - d_{lj}$.

Capita in pratica, abbastanza spesso, che due o più differenze risultino uguali fra loro e maggiori delle altre, di modo che non sia possibile determinare univocamente la massima e quindi il punto ${}_1P$. È allora consigliabile ricorrere a qualche convenzione che permetta di rimuovere tali casi di indeterminatezza. Un espediente semplicissimo consiste nel considerare maggiore fra due differenze uguali, poniamo ${}_1\Delta_{k_2}$ e ${}_2\Delta_{k_2}$ dove ${}_1 < {}_2$ la differenza ${}_1\Delta_{kl}$ se $k < l$ e, viceversa, la differenza ${}_2\Delta_{kl}$ se $k > l$. In altri termini, facendo il confronto fra gli elementi di due righe, fra più differenze uguali si considera maggiore la differenza corrispondente alla colonna che viene prima quando si procede da sinistra a destra, se la riga dei minuendi è posta sopra la riga dei sottraendi. In caso contrario, si considera maggiore la differenza corrispondente alla colonna posta più a destra. È facile vedere che tale convenzione non porta a contraddizioni e non altera la soluzione ottima, se non nel senso di farne prescegliere una sola, anche quando vi sono due o più percorsi ottimi di uguale lunghezza.

Tale dimostrazione si basa sul fatto che il problema non viene ad alterarsi sostanzialmente se immaginiamo sostituite alle distanze d_{ij} per esempio le quantità $d'_{i,j}$ così definite :

$$d'_{ij} = d_{ij} + \xi^{(n-0,5i)} (i-1)+j-i \quad (i < j)$$

$$d'_{ij} = d_{ij} + \xi^{(n-0,5j)} (j-1)+i-j \quad (j > i)$$

dove ξ è una quantità positiva talmente piccola da soddisfare i seguenti requisiti :

1) $n\xi$ è minore dell'elemento più piccolo dell'insieme costituito delle differenze non nulle fra tutte le coppie di percorsi possibili ;

2) le potenze più elevate di ξ sono trascurabili rispetto a quelle di ordine inferiore.

La condizione 1) fa sì che l'introduzione della quantità ξ non alteri la graduatoria dei percorsi di lunghezza diversa ; la condizione 2) conduce appunto alla convenzione sopra descritta.

Usando tale espediente, otteniamo nel nostro esempio, come già accennato sopra, ${}_1P_{6,5} = P_3$; ${}_2P_{6,5} = P_{2j}$; ${}_3P_{6,5} = P_4$

Altre eliminazioni possono essere ottenute immediatamente col ricorso ad una importante proprietà dimostrabile con considerazioni geometriche di carattere elementare, che qui omettiamo. Tale proprietà può essere così enunciata : il percorso ottimo non può contenere nessun tratto che unisca due punti, toccando un altro punto. Tale requisito che integra quello che nel paragrafo precedente è stata designata come proprietà *a*), in quanto stabilisce che nessun punto può essere attraversato, nella soluzione ottima, più di una volta, si traduce nell'eliminazione del tratto $P_k - P_k$ se per un e qualunque si ha : $d_{h,k} - d_{h,l} = d_{k,l}$.

Tornando ancora all'esempio sopra trattato, si vede che in base a tale proprietà il tratto $P_1 - P_4$ può essere eliminato, data la relazione : $d_{1,4} - d_{4,5} = 2-1 = 1 = d_{1,5}$.

Tali eliminazioni possono, ovviamente, essere operate anche e con vantaggio, con mezzi grafici, purché si disponga di una pianta sufficientemente accurata dei punti.

Le sub-eliminazioni e le eliminazioni finora discusse sono probabilmente quelle di più semplice applicazione, in quanto si fondano su criteri assolutamente elementari e del tutto meccanici. Esse possono essere efficacemente messe in evidenza in tabelle a doppia entrata. Si può, per esempio, iscrivere nella casella che si trova all'incrocio della riga *i*-esima con la colonna *j*-esima l'indice *k*, per significare che il tratto $P_i - P_j$ è sub-eliminato da P_k , ossia che il percorso ottimo può contenere il tratto $P_i - P_j$ solo se contiene anche quello $P_j - P_k$. Nel nostro esempio, si giunge così alla tabella N. 5.

TABELLA N. 5

	1	2	3	4	5	6
1	—	3	2,4	5,6		
2		—		3,6		1,4,5
3	2,6		—			1,4,5
4	5,6	1,3	2	—		
5	6	1,3	2,4		—	
6		1,3,5	2,4,5	5		—

I risultati della tabella suddetta possono essere utilmente tradotti in un'altra tabella in cui le eliminazioni vengono indicate con un simbolo (per esempio X) che viene riportato simmetricamente nelle caselle (ij) e (ji) a significare l'eliminazione del tratto $P_i - P_j$. Le sub-eliminazioni, non essendo simmetriche, vi vengono invece segnate come nel tipo di tabella precedentemente descritto. Per il nostro esempio, si veda la tabella N. 6.

TABELLA N. 6

	1	2	3	4	5	6
1	—	3	×	×		
2		—		×	×	×
3	×		—		×	×
4	×	×	2	—		
5	6	×	×		—	
6		×	×	5		—

In base alla tabella N. 6 è facile vedere come in certi casi favorevoli, le eliminazioni di cui si è discusso sopra, possano essere sufficienti per l'individuazione rigorosa del percorso ottimo. Dalla riga (o colonna) 2, risulta che P_2 può essere unito solo con P_1 e P_3 similmente, dalla riga 3 appare chiaro che il percorso ottimo comprende anche il tratto $P_3 - P_4$. Di conseguenza i tratti certi sono dati da $P_1 - P_2 - P_3 - P_4$. Il tratto $P_5 - P_1$ è da considerarsi eliminato, in quanto richiederebbe contemporaneamente il tratto $P_1 - P_6$, il che sarebbe in contrasto col risultato precedentemente raggiunto, secondo cui nel percorso ottimo deve figurare il percorso $P_1 - P_2$. Di conseguenza, P_4 deve essere unito a P_5 e P_6 .

In conclusione, il percorso ottimo (tenuto conto, naturalmente, della convenzione adottata per la scelta fra percorsi di uguale lunghezza) risulta quindi essere dato da $P_1 - P_2 - P_3 - P_4 - P_5 - P_6 - (P_1)$.

3) Finora sono stati passati in rassegna i metodi di eliminazione più semplici ed elementari che si allacciano alle proprietà a) e b) del percorso ottimo e si fondano esclusivamente sul confronto fra coppie di righe della matrice delle distanze.

Altre eliminazioni, e sub-eliminazioni o di carattere leggermente più complesso, si fondano sulla proprietà c) e hanno per oggetto non più tratti fra due punti, ma tratti doppi, o sia tratti che uniscono tre punti.

Per impostare tale procedimento, conviene partire da un esame più approfondito della proprietà c). Supponiamo per maggiore semplicità di simboli, che in un dato problema il percorso ottimo sia costituito da: $P_1 - P_2 - \dots - P_{k-1} - P_k - P_{k+1} - \dots - P_n - (P_1)$. Immaginiamo ora di togliere da tale percorso un punto qualunque, poniamo P_k . Per gli $n-1$ punti rimasti si stabilisce in tal modo il percorso: $P_1 - P_2 - \dots - P_{k-2} - P_{k-1} - P_{k+1} - P_{k+2} - \dots - P_n - (P_1)$.

Notiamo per inciso che tale percorso può essere, ma non è necessariamente, il percorso ottimo fra i punti residui. Immaginiamo ora di inserire nuovamente il punto P_k , tornando in tal modo al problema originario. Dove va inserito questo punto? Evidentemente, là dove il suo inserimento produce il prolungamento minore rispetto al percorso fra gli $n-1$ punti; il prolungamento essendo misurato dalle quantità $(k; q, r) = d_{k,q} + d_{k,r} - d_{q,r}$

quando si suppone che P_k venga inserito fra P_q e P_r . Si deve pertanto avere :

$$(k ; k - 1), k + 1 < (k ; j, j + 1) \quad (j = 1, 2, \dots, n \neq k - 1, \neq k) \quad (9).$$

A maggior ragione sarà quindi verificata, per ogni valore di j , al variare di x , la seguente relazione :

$$(k ; k - 1, k + 1) < \max. (k ; j, x) \quad (x \neq k, \neq j) \quad (10)$$

Tale relazione è ovviamente valida anche per $j = k - 1$ e per $j = k + 1$.

Inoltre, poichè ogni punto, nel percorso dei residui $n - 1$ punti, è unito ad altri due, la quantità $(k ; k - 1, k + 1)$ non può superare la più piccola delle quantità :

$$(k ; j, y_{kj}),$$

al variare di j , dove y_{kj} è il valore di x per il quale la quantità $(k ; j, x)$ assumere il suo secondo massimo. In altri termini si ha :

$$P_{kj} = {}_2P_{k,j} \quad (11)$$

La quantità : $\min. (k ; j, y_{kj})$ costituisce pertanto un « valore critico », che può essere utilizzato per eliminare numerosi tratti doppi. Infatti, da quanto precede, risulta evidente che, comunque scelto il punto P_k , se, per esempio, la quantità $(k ; q, r)$ eccede tale valore critico, nel percorso ottimo non potrà figurare il tratto doppio $P_q - P_k - P_r$.

Anche qui, la convenzione prima introdotta permette di eliminare qualunque ambiguità nel caso che due « prolungamenti » prodotti da uno stesso punto, risultino uguali. In questi casi si può applicare la seguente regola che visualizza i risultati che si ottengono adottando la convenzione suddetta. Si determinino le caselle poste al di sopra la diagonale principale che nella matrice delle distanze corrispondono alle distanze che intervengono nei prolungamenti in questione. Delle caselle che non sono comuni ai due « prolungamenti », si prende quella che è posta sulla riga più alta, o, a parità di riga, quella che si trova sulla colonna più

a sinistra. E precisamente, se la casella in questione è posta sulla colonna o riga corrispondente al punto per il quale si ricerca il valore critico (punto che, chiameremo d'ora in poi punto centrale, in quanto è posto al centro dei tratti doppi che si considerano) il prolungamento del quale fa parte la distanza scritta nella casella è da considerarsi maggiore dell'altro (in quanto la potenza più bassa di ξ appare al minuendo), viceversa, esso è da considerarsi minore quando questo non avviene (poichè allora la potenza più bassa di ξ appare al sottraendo). Così, per tornare al nostro esempio, la quantità $(5 ; 1,2)$ e $(5, 3 ; 4)$ risultano uguali. Si ha infatti :

$$(5 ; 1,2) = d_{1,5} + d_{2,5} - d_{1,2} = 1 + 2 - 2 = 1$$

$$(5 ; 3,4) = d_{3,5} + d_{4,5} - d_{3,4} = 2 + 1 - 2 = 1$$

Secondo la regola enunciata, dovremo poi considerare $(5 ; 1,2)$ più piccolo di $(5 ; 3,4)$. Infatti, la casella determinante è in questo caso la $(1, 2)$ che non si trova nè sulla riga, nè sulla colonna 5. Tale risultato è naturalmente in armonia con la nostra convenzione, si ha infatti :

$$(d_{1,5} + \xi^4) + (d_{2,5} + \xi^8) - (d_{1,2} + \xi) < (d_{3,5} + \xi^{11}) + (d_{4,5} + \xi^{13}) - (d_{3,4} + \xi^{10})$$

analogamente si trova :

$$(1 ; 4,5) > (1 ; 4,6) ; (2 ; 3,5) > (2 ; 3,6) \text{ e così via.}$$

In pratica, tuttavia, i casi di uguaglianza fra prolungamenti sono rilevanti, ai fini del procedimento sommariamente delineato, solo quando si verificano al livello del « valore critico ». In generale, i valori dei vari « prolungamenti » in sè non hanno importanza ; occorre stabilire il « valore critico » e determinare quindi quali sono i « prolungamenti » minori e quali quelli maggiori di tali « valori critici ».

Nel nostro esempio, il procedimento descritto condurrebbe ai risultati esposti nelle tabelle numero 7 - 12.

TABELLA N. 7

Punto centrale: P_1

« Valore critico »:

1,1 (1 = 1; 6,5)

	2	3	4	5	6
2	—	×	×		
3	×	—	×	×	
4	×	×	—	×	×
5		×	×	—	
6			×		—

TABELLA N. 8

Punto centrale: P_2

« Valore critico »:

1 (2; 3,5)

	1	2	3	4	5
1	—		×	×	×
3		—	×		
4	×	×	—	×	×
5	×		×	—	×
6	×		×	×	—

TABELLA N. 9

Punto centrale: P_3

« Valore critico »:

1 (3; 2,6)

	1	2	4	5	6
1	—	×	×	×	×
2	×	—			
4	×		—	×	×
5	×		×	—	×
6	×		×	×	—

TABELLA N. 10

Punto centrale: P_4

« Valore critico »:

1,1 (4; 6,5)

	1	2	3	5	6
1	—	×	×	×	×
2	×	—	×	×	
3	×	×	—		
5	×	×		—	
6	×				—

TABELLA N. 11

Punto centrale: P_5

« Valore critico »:

0,9 (5; 4,6)

	1	2	3	4	6
1	—	×			×
1	×	—	×		
3		×	—	×	
4			×	—	
6	×				—

TABELLA N. 12

Punto centrale: P_6

« Valore critico »:

3,1 (6; 1,3)

	2	3	4	5
1	—	×		
2	×	—	×	×
3	×	—	×	×
4		×	×	—
5	×	×	×	—

Alle eliminazioni iniziali, cui si perviene con tale procedimento, possono aggiungersene altre, a mano a mano che, coi mezzi finora descritti, o con altri metodi, aumentano le nostre cognizioni sulla soluzione ottima, oppure, il che è la stessa cosa, a mano a mano che si riduce l'insieme dei tratti che potrebbero figurare nel percorso ottimo.

Così, nel nostro esempio, se si suppone eliminato il tratto 4-6, il « valore critico » per il punto centrale P_1 diventa uguale a 0,9, corrispondente al prolungamento (1; 6,3). Dal ragionamento che ci ha condotto alla nozione di « valore critico » risulta, infatti, che tale valore deve essere inferiore a quello corrispondente al prolungamento maggiore, in ogni riga, o colonna, tenuto però conto soltanto di quelle coppie di punti, che possono trovarsi unite da tratti diretti nel percorso ottimo.

Con tale spostamento del « valore critico », nel caso nostro, si eliminerebbero i tratti 2-1-5 e 5-1-6 e si dimostrerebbe, quindi, la necessità che nel percorso ottimo figuri il tratto 1-6.

Questo esempio permette di intravedere come le eliminazioni di tratti doppi ora esaminate, e quelle di tratti semplici, prima

discusse, possono utilmente integrarsi a vicenda. Tale possibilità, infatti, è stata da noi sperimentata con successo in alcuni problemi di medie dimensioni.

La portata del procedimento di eliminazione fondato sul « valore critico » non si limita al caso più semplice, finora esaminato. Si supponga infatti, di aver stabilito che nel percorso ottimo figuri il tratto $P_i - P_j$. Si può allora ripetere esattamente il ragionamento fatto prima, purchè si sostituisca al « punto centrale » P , la « coppia centrale » $P_i - P_j$ e purchè si definiscano nel modo seguente i « prolungamenti » rispetto a tale coppia centrale :

$$\{i, j; p, q\} = m_i n (d_{p,i} + d_{q,j} - d_{p,q}; d_{p,j} + d_{q,i} - d_{p,q})$$

$$(p, q = 1, 2, \dots, n; \neq i, j)$$

Lo stesso procedimento si applica ugualmente se P_i e P_j sono i due punti estremi di un tratto composto, comprendente più punti, di cui sia stata accertata la presenza nel percorso ottimo. In tal caso, naturalmente, non si terrà però conto dei punti intermedi.

Data la relazione :

$$d_{p,i} + d_{q,j} - d_{p,q} \geq d_{p,j} + d_{q,i} - d_{p,q} \text{ se}$$

$$d_{p,i} - d_{q,i} \geq d_{p,j} - d_{q,j},$$

si può procedere come segue per individuare la successione dei punti corrispondente ai « prolungamenti » : si calcolano anzitutto tutte le differenze fra i termini corrispondenti delle righe i -esima e j -esima della matrice delle distanze e si ordinano tali differenze in graduatoria decrescente. Sia k_1, k_2, k_3, \dots la successione in tal modo determinata, delle colonne (escluse quelle corrispondenti agli eventuali punti intermedi, cui si è fatto cenno dianzi).

In una tabella a doppia entrata, con le righe e le colonne intestate a tali indici, si pongono allora simboli (per esempio X) che denotino eliminazione, in tutte le caselle della *riga* k_1 . Si procede poi allo stesso modo con la *riga* k_2 , eccettuando la casella posta nella colonna k_1 . Poi si passa alla *riga* k_3 , nella cui casella

va posto lo stesso simbolo, eccezion fatta per le colonne k_1 e k_2 . Si continua così fino al penultimo degli indici (L'ultima riga rimasta non dovrà, ovviamente, contenere simboli di eliminazione). Tale tabella che, per distinguerla da quella contenente i « prolungamenti » veri e propri, chiameremo « ausiliaria », va letta come segue: se nella casella (h, l) si trova collocato un simbolo di eliminazione, il « prolungamento » relativo ai punti P_h, P_l , sempre rispetto ai punti centrali P_i e P_j , si riferisce ai tratti $P_h - P_j - P_i - P_l$, e non ai tratti $P_h - P_i - P_j - P_l$, in base alla relazione: $d_{h,i} + d_{i,l} < d_{h,j} + d_{j,l}$. In altri termini, i tratti da prendere in considerazione sono quelli relativi alle caselle non contrassegnate, le quali vanno interpretate nel senso che, se il primo estremo è il punto corrispondente alla riga, il secondo è il punto P_i , il terzo è il punto P_j (si trascurano qui gli eventuali punti intermedi, di cui, come si è detto, non si tiene conto), e l'altro estremo è dato dal punto corrispondente alla colonna.

Illustriamo tale procedimento, tornando al nostro esempio. Come si è visto, è risultato dimostrato che il percorso ottimo comprende il tratto $P_2 - P_3$. Per costruire la tabella ausiliaria, rispetto alla coppia centrale P_2, P_3 si calcolano anzitutto le differenze fra gli elementi corrispondenti alle righe 2 e 3 della matrice delle distanze, e si graduano tali differenze, come riportato qui di seguito:

	1	4	5	6
2	2	2,6	2	3,7
3	2,6	2	2	3,7
Differenza.	-0,6	0,6	0	0
Numero di graduatoria	4	1	2	3

A tale graduatoria corrisponde, secondo il procedimento sopra esposto, la tabella n. 13 che rappresenta la tabella ausiliaria, relativa alla coppia centrale $P_2 - P_3$.

TABELLA N. 13

Punti centrali : $P_2 - P_3$
tabella ausiliaria

	1	4	5	6
1	—			
4	×	—	×	×
5	×		—	×
6	×			—

TABELLA N. 14

Punti centrali : $P_2 - P_3$
Prolungamenti

	1	4	5	6
1	—	2	3	3,7
4	2	—	3	3,7
5	3	3	—	3,8
6	3,7	3,7	3,8	—

Nella tabella n. 14 sono stati messi in evidenza i prolungamenti relativi alla coppia $P_2 - P_3$, in base alla tabella ausiliaria N. 13. Ne risulta che il valore critico, corrispondente al prolungamento (2, 3; 4, 5) che come si vedrà meglio in seguito, secondo la convenzione adottata deve essere considerato inferiore al prolungamento (2, 3; 1, 5) è pari a 3. Sono quindi da eliminare i tratti aventi per estremi i punti $P_1 - P_5$, $P_1 - P_6$, $P_4 - P_6$, $P_5 - P_6$. Tali eliminazioni vanno riportate nella tabella ausiliaria che assume pertanto l'aspetto riportato nella tabella a N. 15.

TABELLA N. 15

	1	4	5	6
1	—		×	×
4	×	—	×	×
5	×		—	×
6	×	×	×	—

Dalla tabella N. 15 risulta stabilito il tratto $P_3 - P_4$, in quanto tutte le caselle prive di simbolo di eliminazione (X) si trovano nella colonna corrispondente a P_4 . Viceversa, poichè tale simbolo è posto in tutte le caselle della colonna corrispondente a P_6 risulta eliminato il tratto $P_3 - P_6$. Risultando stabilito il tratto

$P_3 - P_4$ in aggiunta al $P_2 - P_3$, il tratto $P_2 - P_4$ sarebbe naturalmente eliminato, anche se tale eliminazione non fosse già avvenuta precedentemente. È chiaro che, così come le eliminazioni risultanti dal confronto dei vari « prolungamenti » col valore critico vanno riportate sulla tabella ausiliaria, su questa vanno registrate anche le altre eliminazioni, individuate comunque. Tali eliminazioni vanno poi riportate anche sulla tabella dei prolungamenti. Così nel nostro esempio, se si riuscisse in qualche modo a dimostrare che il tratto $P_2 - P_5$ non può figurare nel percorso ottimo, il segno X andrebbe posto nella casella (4,5) della tabella N. 15 e l'eliminazione andrebbe registrata ugualmente nelle caselle (4,5) e (5,4) della tabella N. 14, accanto a quelle riguardanti le caselle contenenti valori superiori al « valore critico ».

Tale osservazione può assumere una certa importanza in problemi di più vaste dimensioni, anche perchè permette in molti casi di ridurre i calcoli. Non è infatti necessario calcolare i « prolungamenti » relativi ai tratti eliminati in tal modo. Il « valore critico » può essere determinato senza che si conoscano tutti i « prolungamenti ». Esso è dato dalla più piccola delle seguenti quantità al variare di h :

$d_{i,h} + d_{i,h'(i)} - d_{h'(i),h}$; $d_{i,h} + d_{j,h'(j)} - d_{h'(j),h}$ dove P_i e P_j sono i « punti centrali » e dove $P_{h'(i)} = {}_2P_{i,h}$; $P_{h'(j)} = {}_2P_{j,h}$. In tal modo, si possono trascurare completamente quei punti di cui si è accertato che non possono essere direttamente collegati con nessuno dei due punti centrali nel percorso ottimo.

Tornando al nostro esempio, dopo aver stabilito che il percorso ottimo contiene il duplice tratto $P_2 - P_3 - P_4$, si può procedere allo stesso modo prendendo come punti centrali i punti estremi di tale tratto cioè P_2 e P_4 , e trascurando il punto intermedio, P_3 . La tabella ausiliaria e la tabella dei prolungamenti sono riportati rispettivamente come tabella N. 16 e N. 17.

TABELLA N. 16

	I	5	6
I	—		
5	×	—	
6	×	×	—

TABELLA N. 17

	I	5	6
I	—	2	2
5	2	—	2, I
6	2	2, I	—

Dalle tabelle n. 16 e n. 17 risulta anzitutto che il percorso ottimo non può contenere insieme i tratti $P_5 - P_1$ e $P_4 - P_6$. E' questo infatti il senso del simbolo di eliminazione che, secondo i criteri precedentemente chiariti, va posto nella casella (5,6), alla quale corrisponde il « prolungamento » maggiore.

Non rimane ormai che stabilire, in base alla convenzione adottata per la scelta fra tratti di uguale lunghezza, se vada eliminato il tratto avente per estremi i punti P_1 e P_5 o quello avente per estremi i punti P_1 e P_6 . A tale scopo ci si può servire dei tratti doppi. Basta, cioè, verificare a quali caselle di una matrice triangolare corrispondono i tratti compresi nei subpercorsi in questione. Nel nostro caso, prendendo nuovamente in considerazione la parte di una matrice quadrata posta al di sopra della diagonale principale, vediamo dalla tabella n. 16 che si tratta, rispettivamente, delle caselle (1,2), (4,5), (1,5) e (1,2), (4,6), (1,6). Le caselle poste più in alto, fra quelle che non sono comuni ai due percorsi, sono la (1,5) e la (1,6). Di queste, quella posta più a sinistra è la (1,5). È quindi su questa che dobbiamo soffermarci per stabilire quale dei due prolungamenti debba considerarsi maggiore.

Poichè nè la riga, nè la colonna sono intestati a « punti centrali » (nel nostro caso : i punti P_2 e P_4), la casella in questione corrisponde al sottraendo dell'espressione che dà il « prolungamento » ; di conseguenza il « prolungamento » nel quale figura la distanza in questione, è da considerarsi minore dell'altro. Anche per questa via, in conclusione, si è dimostrato che il percorso più breve è dato da : $P_1 - P_2 - P_3 - P_4 - P_5 - P_6 - (P_1)$.

4) In problemi di piccole dimensioni, come quello esaminato a titolo di esempio, e anche, talvolta, in problemi di maggior mole, solo nei casi più fortunati, però, i procedimenti sopra delineati sono di per se stessi sufficienti e permettono di giungere con una certa facilità, in modo del tutto rigoroso, alla soluzione ottima.

Nei problemi di medie e di grandi dimensioni, tuttavia, tali procedimenti che, come si è detto, si fondano su condizioni necessarie, ma non sufficienti, pur risultando utili, non bastano, in generale, per individuare in modo univoco il percorso più breve o per dimostrare che un determinato percorso sia ottimo. In tali casi occorre fare ricorso anche ad altri procedimenti, fra

cui assume un ruolo di primo piano l'applicazione di alcuni principi della programmazione lineare, come è già stato brevemente accennato nel primo paragrafo.

Come accennato prima, il modo più semplice e, forse, anche più efficace, di applicare al problema del « Commesso viaggiatore » tali principi, consiste nel rendere massima la funzione :

$$\sum_{i=1}^n u_i \quad (12)$$

subordinatamente alle condizioni :

$$d_{i,j} - u_i - u_j \geq 0 \quad (i, j = 1, 2 \dots n ; i \neq j) \quad (13)$$

In tale forma, il problema non rientra a rigore, nella programmazione lineare in senso stretto, in quanto le variabili u_i non sono necessariamente non negative. Si tratta, però, di una divergenza apparente ; infatti, per ottenere la piena applicabilità dello schema della programmazione lineare, basta introdurre nella (13) le variabili non negative $\delta_{i,j}$ già definite. Le disuguaglianze (13) diventano allora equazioni :

$$d_{i,j} - u_i - u_j = \delta_{i,j} \quad (i, j = 1, 2, \dots, n ; i \neq j) \quad (13,a)$$

Si possono allora, con l'ausilio della (13,a), eliminare le u_i esprimendole per mezzo di n variabili $\delta_{i,j}$ opportunamente scelte, e sostituire tali espressioni nella funzione (12), come anche nelle restanti equazioni (13,a). In tal modo si è ricondotti ad un problema di programmazione lineare, nell'accezione comune, comprendente $\binom{n-1}{2}$ variabili non negative, soggette a $\binom{n-1}{2} - n$ condizioni. Come è noto, almeno una soluzione ottima di tale problema è costituita da una « soluzione base », ossia, nel nostro caso, da una di quelle soluzioni che soddisfano le condizioni poste e in cui almeno n delle variabili $\delta_{i,j}$ sono uguali a 0, mentre quelle diverse da 0 rappresentano un insieme di variabili indipendenti.

Il problema così posto può essere considerato come uno dei più semplici fra quelli rientranti nel vasto campo della « pro-

grammazione lineare», in quanto sia la forma della funzione da ottimizzare, sia quella delle equazioni vincolanti (in cui, almeno inizialmente, non compaiono coefficienti diversi da -1), è la più elementare possibile. Di conseguenza, anche per tale tipo di problema come per quelli strettamente correlati, ma alquanto più generali, che si intitolano « problema dei trasporti » e « problema della distribuzione del personale » (« personal assignments problem »), sono stati escogitati alcuni procedimenti tendenti a consentire il raggiungimento di una soluzione, per lo meno approssimata, con calcoli notevolmente ridotti rispetto a quelli richiesti dai metodi tradizionale.

Mentre si rinviano i lettori che desiderassero approfondire l'argomento ai lavori citati nella bibliografia (si vedano in particolare gli articoli di Dantzig e Croes, per quanto riguarda i metodi attualmente più usati, verrà esposto qui di seguito alcuni nuovi procedimenti alternativi che in diversi casi si sono dimostrati abbastanza efficaci. Tali procedimenti hanno in comune con quelli già noti il fatto di fare perno, piuttosto sulle quantità u_i (che qualche autore definisce « potenziali »), che sulle variabili.

Il primo di questi procedimenti si fonda sul criterio di aumentare gradualmente i « potenziali » u_i , facendo però sì che, ove possibile, nella risultante tabella delle distanze modificate, δ_{ij} , nella riga (colonna) alla quale corrisponde il potenziale aumentato per ultimo, non si trovi, ogni volta, più di un elemento che rappresenti il minimo assoluto di un'altra colonna (riga). Tale criterio si può giustificare empiricamente con la considerazione che, quando in una riga, poniamo nella riga h , si trovano i minimi di due colonne, poniamo le colonne k e l , un aumento di u_h , facendo ulteriormente decrescere tali minimi, verrebbe con ogni probabilità a ridurre il valore che altrimenti potrebbero assumere u_k e u_l e si risolverebbe quindi in definitiva in una diminuzione della somma di tutte le u_i , rispetto al massimo che quella somma può attingere.

Il procedimento prende le mosse dalla tabella delle distanze effettive, quando cioè tutti i « potenziali » sono nulli, e prosegue con l'aumento del valore di un potenziale alla volta, accompagnato talvolta, per non violare la condizione di non-negatività delle variabili δ_{ij} dalla diminuzione di un altro potenziale. Più precisamente, in qualunque stadio del procedimento, si prende

in considerazione, nella tabella delle δ_{ij} , la riga (la colonna) avente il più grande dei minimi, per righe (colonne). Tale scelta garantisce infatti automaticamente, che la riga in questione non contenga minimi di colonne (righe). Si fanno poi le differenze fra le δ_{ij} della riga (colonna) prescelta e i minimi delle colonne corrispondenti. La seconda di queste differenze, in ordine crescente, rappresenta la quantità da aggiungere al potenziale, corrispondente alla riga. Dopo tale incremento, la riga (colonna) in questione conterrà, in linea generale, un minimo assoluto per colonna (riga) e un altro minimo eguagliato; vi possono però essere anche altri minimi eguagliati. Può avvenire, come già si è accennato, che il minimo assoluto di una colonna (riga) risulti negativo; in tal caso occorre diminuire il potenziale corrispondente alla linea in questione, in modo da annullare la δ risultata negativa. In simboli, se dopo l'aumento, poniamo, di u_h , risulta:

$\delta_{h,h} = d_{h,h} - u_h - u_h < 0$, occorre ridurre subito dopo, in pratica anche temporaneamente, il potenziale u_h della quantità: $u_h + u_h - d_{h,h}$.

Una certa complicazione sorge quando diverse righe (colonne) hanno minimi uguali, maggiori di quelle delle altre righe (colonne). In tal caso, si accrescono i le potenziali corrispondenti a righe (colonne) — se ve ne sono — in cui non vi sono minimi di colonne (righe) e, subordinatamente, quelle in cui figura solo uno di tali minimi.

Il procedimento deve essere proseguito, se possibile, fino a quando in ogni riga (colonna) vi è almeno una δ uguale a zero. Talvolta, però, occorre arrestarsi prima, in quanto in un certo numero di righe (colonne) figurano due o più minimi di colonna (riga), diversi da zero, mentre nelle altre righe (colonne) figurano degli zeri e quindi le regole annunciate non risultano più applicabili. In tali casi, però, è spesso possibile, come si vedrà più avanti, determinare, con facili calcoli, i valori che devono essere attribuiti ai potenziali perchè tutti i minimi di riga (colonna) si annullino.

In pratica, converrà disporre i calcoli in modo da mettere in evidenza, in ogni fase, oltre alle variabili δ , anche, in un'apposita riga (colonna) i minimi di colonna o riga di tali variabili, come è fatto nelle tabelle n. 18 — n. 22 che si riferiscono sempre al nostro esempio.

TABELLA N. 18

i/j	1	2	3	4	5	6	u
1	—	2	2,6	2	1	2	0
2	2	—	1	2,6	2	3,7	0
3	2,6	1	—	2	2	3,7	0
4	2	2,6	2	—	1	2	0
5	2	2	1	1	—	1,9	0
6	2	3,7	3,7	2	1,9	—	0
minimi	1	1	1	1	1	1,9	

TABELLA N. 19

i/j	1	2	3	4	5	6	u
1	—	2	2,6	2	1	1	0
2	2	—	1	2,6	2	2,7	0
3	2,6	1	—	2	2	2,7	0
4	2	2,6	2	—	1	1	0
5	1	2	2	1	—	0,9	0
6	1	2,7	2,7	1	0,9	—	1
minimi	1	1	1	1	0,9	0,9	

TABELLA N. 20

i/j	1	2	3	4	5	6	u
1	—	1,9	2,5	1,9	0,9	0,9	0,1
2	1,9	—	1	2,6	2	2,7	0
3	2,5	1	—	2	2	2,7	0
4	1,9	2,6	2	—	1	1	0
5	0,9	2	2	1	—	0,9	0
6	0,9	2,7	2,7	1	0,9	—	1
Min.	0,9	1	1	1	0,9	0,9	

TABELLA N. 21

i/j	1	2	3	4	5	6	u
1	—	1,9	2,5	1,8	0,9	0,9	0,1
2	1,9	—	1	2,5	2	2,7	0
3	2,5	1	—	1,9	2	2,7	0
4	1,8	2,5	1,9	—	0,9	0,9	0,1
5	0,9	2	2	0,9	—	0,9	0
6	0,9	2,7	2,7	0,9	0,9	—	1
Min.	0,9	1	1	0,9	0,9	0,9	

TABELLA N. 22

i/j	1	2	3	4	5	6	u
1	—	0,9	2,5	1,8	0,9	0,9	0,1
2	0,9	—	0	1,5	1	1,7	1
3	2,5	0	—	1,9	2	2,7	0
4	1,8	1,5	1,9	—	0,9	0,9	0,1
5	0,9	1	2	0,9	—	0,9	0
6	0,9	1,7	2,7	0,9	0,9	—	1
Min.	0,9	0	0	0,9	0,9	0,9	

Le tabelle n. 18 e n. 19 appaiono sufficienti per illustrare il procedimento esposto. La tabella n. 18, che è la tabella iniziale, non è altro che la tabella delle distanze (n. 1) alla quale sono state aggiunte una riga in cui sono riportati tutti i minimi di colonna e una colonna in cui sono riportati i valori di « potenziali » che in quella fase iniziale sono naturalmente tutti nulli. Dalla riga dei minimi di colonna si vede che il maggiore di tali minimi corrisponde alla colonna 6. Di conseguenza, occorre anzitutto accrescere il valore di u_6 , e precisamente di una quantità che si determina calcolando le differenze fra gli elementi della riga 6 e quelli corrispondenti dell'ultima riga che appunto contiene i minimi di colonna. La minore di queste differenze è pari a 0,9 (in corrispondenza della colonna 5); vengono poi le differenze relative alle colonne 1 e 4 che sono entrambe uguali a 1 e questo secondo minimo rappresenta appunto la quantità di cui deve essere accresciuta u_6 . Il risultato di tale operazione è riportato nella tabella n. 19, nell'ultima colonna della quale, mentre gli altri « potenziali » figurano sempre col valore 0, u_6 figura col valore 1 ($= 0 + 1$). Di conseguenza, gli elementi della riga e colonna 6 della tabella precedente vengono tutti ridotti

di 1. Nella riga dei minimi vengono modificati quelli corrispondenti alle colonne 5 e 6. Tutti gli altri elementi della tabella n. 18, invece, ricompaiono senza modificazione nella tabella n. 19. Per proseguire ora il procedimento, occorre anzitutto determinare quale « potenziale » vada accresciuto. All'uopo non basta esaminare l'ultima riga, poichè da questa risulta che ben 4 colonne hanno dei minimi eguali a 1, superiori a quelli delle altre 2 colonne. Consultando la tabella, però, si vede che le righe 1 e 4 a differenza delle righe 2 e 3 non contengono minimi di colonne. Possiamo quindi, indifferentemente, accrescere u_1 o u_4 . Nel caso in questione, non vi sarebbe nessuna differenza sostanziale; si giungerebbe comunque, anche se fosse stato prima aumentato u_4 anzichè u_1 , come è stato invece fatto, alla tabella N. 21, la quale presenta poi una analoga indeterminatezza nei riguardi di u_2 e u_3 . Se si aumenta u_2 si giunge alla situazione rispecchiata nella tabella n. 22. Le colonne (righe) 2 e 3 contengono un elemento nullo, le altre hanno minimi uguali, pari a 0,9, e per di più ogni riga (colonna) contiene due minimi di colonna (riga). Più precisamente, le righe 1 e 4 contengono minimi delle colonne 5 e 6, la riga (colonna) 5 contiene minimi delle colonne (righe) 1,4 e 6 e la riga (colonna) 6, infine, contiene minimi delle colonne (righe) 1,4 e 5. Non sono quindi più applicabili le regole fin qui seguite. Si vede però facilmente che basta accrescere di 0,45 i valori di u_1 , u_4 , u_5 e u_6 per ridurre a 0 contemporaneamente tutti i minimi suddetti. Così facendo si giunge alla tabella n. 23.

È interessante osservare che la tabella n. 23, nonostante le apparenze, non rappresenta una « soluzione base », come è dimostrato dal fatto che il numero degli elementi nulli può essere aumentato di uno, (per esempio, aumentando u_2 e diminuendo u_3), senza che venga alterato il valore delle variabili nulle che costituiscono l'insieme delle variabili non di base. Difatti, le variabili non-nulle corrispondenti a tali tabelle non sono tutte indipendenti l'una dall'altra. Aumentando u_2 , in modo da annullare $\delta_{1,2}$ e diminuendo u_3 , in modo da lasciare nulla la variabile $\delta_{2,3}$, si giunge alla soluzione-base riportata nella tabella n. 4. Aumentando invece u_3 e diminuendo u_1 , si giunge alla soluzione-base riportata nella tabella n. 24. Entrambe queste soluzioni-base sono « degenerate »; esse contengono cioè, ognuna, una variabile-base nulla. Lo stesso dicasi delle so-

TABELLA N. 23

	1	2	3	4	5	6	u_i
1	—	0,45	2,05	0,9	0	0	0,55
2	0,45	—	0	1,05	0,55	1,25	1
3	2,05	0	—	1,45	1,55	2,25	0
4	0,9	1,05	1,45	—	0	0	0,55
5	0	0,55	1,55	0	—	0	0,45
6	0	1,25	2,25	0	0	—	1,45
minimi	0	0	0	0	0	0	

TABELLA N. 24

i/j	1	2	3	4	5	6	u_i
1	—	1,9	0,6	0,9	0	0	0,55
2	1,9	—	0	1,5	1	1,7	—0,45
3	0,6	0	—	0	0,1	0,8	1,45
4	0,9	1,5	0	—	0	0	0,55
5	0	1	0,1	0	—	0	0,45
6	0	1,7	0,8	0	0	—	1,45
minimi	0	0	0	0	0	0	

luzioni riportate nelle tabelle n. 2 e n. 3. Con le tabelle citate si esaurisce l'insieme delle soluzioni — base ottime; ogni altra soluzione ottima, per un noto teorema della programmazione lineare, deve essere esprimibile come combinazione lineare di tali soluzioni — base. Così, per esempio, la tabella n. 23, che rappresenta ugualmente una soluzione ottima, può essere ottenuta dalla media ponderata delle tabelle n. 4 e n. 24 se a queste si attribuiscono rispettivamente i pesi $29/38$ e $9/38$.

La dimostrazione che ognuna delle tabelle citate corrisponde ad una soluzione ottima, è nel caso specifico abbastanza semplice. Prendendo, per esempio la tabella n. 23 è facile osservare che, date le condizioni imposte, la somma $\sum_{i=1}^6 u_i$, non può superare ovviamente la quantità: $d_{1,5} + d_{2,3} + d_{4,6} = 4$.

Poichè i valori attribuiti ai potenziali u_i sono tali da uguagliare appunto tale quantità (come si vede subito, anche senza fare la somma, per fatto che le d_{ij} corrispondenti a quelle distanze sono nulli), la tabella n. 23 rappresenta una soluzione ottima. La questione del criterio per determinare se una soluzione data è ottima, verrà comunque ripresa più sotto, su un piano più generale.

Il procedimento ora esposto dà in generale buoni risultati, nel senso che, nelle applicazioni pratiche che ne sono state fatte, esso ha quasi sempre permesso di individuare con esattezza le variabili-base di una soluzione-base ottima. D'altro lato, però, non sempre tale procedimento risulta rapido come sarebbe desiderabile, essendo congegnato in modo da condurre ad una soluzione ottima attraverso un numero piuttosto elevato di operazioni comportanti ciascuna un miglioramento di portata limitata.

In certi casi, soprattutto quando si ritiene di potersi accontentare di una soluzione approssimata del problema di programmazione lineare che stiamo trattando, può essere più consigliabile il procedimento che ora esporremo. Tale procedimento si differenzia da quello precedente perchè in ogni stadio al potenziale da aumentare viene aggiunta una quantità uguale al secondo minimo della riga (colonna) corrispondente. Ciò naturalmente comporta in pari tempo una diminuzione del potenziale corrispondente al minimo della riga (colonna) del primo potenziale, in modo che risulti nullo anche tale minimo. Questo procedimento è illustrato nelle tabelle n. 25 e n. 26.

TABELLA N. 25

i/j	1	2	3	4	5	6	u_i
1	—	2	2,6	2	1,1	0	0
2	2	—	1	2,6	2,1	1,7	0
3	2,6	1	—	2	2,1	1,7	0
4	2	2,6	2	—	1,1	0	0
5	1,1	2,1	2,1	1,1	—	0	—0,1
6	0	1,7	1,7	0	0	—	2
Min.	0	1	1	0	0	0	

TABELLA N. 26

i/j	1	2	3	4	5	6	u_i
1	—	0,3	3,3	2	1,1	0	0
2	0,3	—	0	0,9	0,4	0	1,7
3	3,3	0	—	2,7	2,8	2,4	—0,7
4	2	0,9	2,7	—	1,1	0	0
5	1,1	0,4	2,8	1,1	—	0	—0,1
6	0	0	2,4	0	0	—	2
Min.	0	0	0	0	0	0	

La tabella iniziale coincide naturalmente con la tabella n. 18. Nella tabella n. 26 vi è un elemento nullo in ogni riga (colonna). Volendo giungere ad una soluzione migliore, conviene anzitutto ricercare se è possibile aumentare la somma delle u_i lasciando inalterate tutte le δ_{ij} uguali a 0. A tale scopo è utile esaminare l'effetto delle modificazioni che si determinano aumentando il potenziale corrispondente ad una riga (colonna) contenente un solo elemento nullo. Cominciando, per esempio, dalla prima riga, si trova che ad un aumento di u_1 corrisponde una diminuzione, di pari ammontare, di u_6 (poichè $\delta_{1,6} = 0$); a questa corrisponde poi un analogo aumento di u_2, u_4 e u_5 , mentre l'aumento di u_2 determina un'uguale diminuzione di u_3 . Tale concatenazione può essere messa in evidenza, per esempio, dal seguente schema :

						5
						4
		6				2
$\delta_{ij} = 0$	6	3	2	6	6	1
<hr/>						
Potenziale :	1	2	3	4	5	6
	+	+	—	+	+	—

In tale schema, sopra ogni indice dei potenziali sono riportati gli indici corrispondenti alle colonne (righe) degli elementi nulli della stessa riga (colonna). Quando un segno (+ o —) viene posto sotto un numero, il segno contrario va posto sotto i numeri uguali a quelli sovrastanti il primo e così via. Poichè il numero dei segni positivi supera nel nostro caso quello dei segni negativi, conviene operare la modificazione sinteticamente indicata dai simboli ; se il numero dei segni negativi fosse stato superiore a quello dei segni positivi, sarebbe stato conveniente effettuare l'operazione a segni invertiti, (cioè aumentando i potenziali contrassegnati col segno — e diminuando quelli contrassegnati col segno +). Se il numero dei segni + fosse uguale a quello dei —, la somma complessiva dei potenziali rimarrebbe ovviamente inalterata sia con l'operazione nei termini indicati dai simboli, sia con quella opposta ; si avrebbe solo il vantaggio di giungere ad una soluzione-base.

Tornando al nostro esempio, si tratta di dare il massimo incremento possibile a u_1, u_2, u_4 e u_5 . Dalla tabella n. 26 si desume che tale incremento non può essere maggiore di 0,15, poichè altrimenti $\delta_{1,2}$ diventerebbe negativa. Nella tabella n. 27 sono riportati i valori risultanti da tale modificazione:

TABELLA N. 27

i, j	1	2	3	4	5	6	n_i
1	—	0	3,3	1,7	0,8	0	0,15
2	0	—	0	0,6	9,1	0	1,85
3	3,3	0	—	2,7	2,8	2,7	-0,85
4	1,7	0,6	2,7	—	0,8	0	0,15
5	0,8	0,1	2,8	0,8	—	0	0,85
6	0	0	2,7	0	0	—	1,85
min.	0	0	0	0	0	0	

Con la tabella n. 27 siamo giunti ad una soluzione-base. È noto dalla teoria della programmazione lineare, che si può giungere ad una soluzione migliore, nel caso che ve ne siano, semplicemente sostituendo ad una variabile-base una variabile non di base, cioè aumentando il valore di una variabile nulla, lasciando invariate le altre, fino ad annullare una variabile diversa da 0. Nel caso specifico, in altre parole, si tratta di modificare i valori di tutti, o alcuni potenziali, in modo però da lasciare invariato δ_{ij} nulle, eccettuata una, che diventerà positiva. Il problema si riduce quindi all'indagine tendente a determinare se sia possibile ottenere una soluzione migliore facendo crescere ora una ora l'altra delle variabili nulle. Tale indagine può essere schematizzata,

in modo del tutto analogo a quello adottato in precedenza, secondo il quadro di cui alla tabella n. 28.

TABELLA N. 28

						5	
		6				4	
	6	3				2	
	2	1	2	6	6	1	
Potenziati	u_1	u_2	u_3	u_4	u_5	u_6	Variabili non di base
	—	—	+	—	—	+	1,2
	—	+	—	+	+	—	1,6
			—				2,3
	+	—	+	+	+	—	2,6
				—			4,6
					—		5,6

Tale schema va letto in questo modo : se si porta nella base, per esempio, $\delta_{1,2}$, cioè se si attribuisce un valore positivo a questa variabile, in modo tale da annullare un'altra variabile, occorre ridurre di una determinata quantità i potenziali u_1 , u_2 , u_4 e u_5 e aumentare della stessa quantità i potenziali u_3 e u_6 . Il risultato complessivo sarebbe quindi negativo. Ugualmente dannosa sarebbe la introduzione nella base di $\delta_{2,3}$, $\delta_{4,6}$ e $\delta_{5,6}$. È invece indifferente introdurre, o no, nella base $\delta_{1,6}$, mentre risulta vantaggioso introdurvi $\delta_{2,5}$ il che comporta la riduzione di 2 potenziali (u_2 e u_6) e l'aumento di 4 (u_1 , u_3 , u_4 , u_5).

Osserviamo per inciso che con tale operazione si giunge alla soluzione ottima di cui alla tabella n. 4.

La costruzione della tavola sopra riportata, è abbastanza semplice : in corrispondenza delle variabili non di base $\delta_{i,j}$ si pone il segno — nella colonna intestata a u_i , si pone poi il segno + nelle colonne intestate a potenziali corrispondenti agli indici *diversi da j*, che sono segnati nella stessa colonna sopra l'intestazione ; si pone poi il segno — nelle colonne corrispondenti agli indici segnati sopra le intestazioni delle colonne relative a questi potenziali e così via. Può tuttavia accadere che, partendo dalla colonna intestata a u_i si giunga ad un certo punto ad una contraddizione, nel senso che in una stessa casella dovrebbe essere

posto sia il segno $+$, sia il segno $-$; occorre allora rifare il procedimento, partendo dalla colonna intestata a u_j . Se per esempio, nella riga corrispondente a $\delta_{2,3}$, si cominciasse col porre il segno $-$ nella colonna intestata a u_2 , bisognerebbe porre successivamente il segno $+$ nelle colonne intestate a u_1 e u_6 , ma poichè sopra u_1 si trova segnato il numero 6, ad indicare la presenza della variabile nulla $\delta_{1,6}$, al $+$ nella colonna u_1 dovrebbe corrispondere un $-$ nella colonna u_6 . Si elimina questa contraddizione partendo invece dalla colonna u_3 .

Dalla tabella appare chiaramente che molti dei potenziali possono essere considerati in coppie, neutralizzandosi a vicenda. Più precisamente, se δ_{ij} è una variabile non di base, ad ogni aumento di u_i corrisponde una diminuzione di u_j , a meno che non si tratti di modificazione tendente a portare nella base appunto δ_{ij} . Inoltre, se vi sono due variabili non di base corrispondenti a caselle di una stessa riga (colonna), i potenziali relativi alle due colonne (righe) o non varieranno affatto, o varieranno nello stesso senso, e cioè in senso opposto al potenziale corrispondente alla riga (colonna), salvo che non si tratti di portare nella base una delle variabili in questione. In altri termini, se $\delta_{h,j}$, e $\delta_{i,j}$ sono variabili non di base, u_h e u_i se variano, variano nello stesso senso e in senso opposto a u_j , a meno che non si tratti di portare nella base proprio $\delta_{h,j}$ o $\delta_{i,j}$.

Appare pertanto naturale, al fine di abbreviare il procedimento per indagare se è possibile migliorare una data soluzione-base, dividere i potenziali in due gruppi tali che ad ogni $\delta_{i,j}$ non di base corrisponda un u_i in un gruppo e un u_j in un altro, per accoppiare il maggior numero possibile di potenziali in modo da neutralizzarli a vicenda. Occorre però osservare che non per tutte le variabili non di base si può operare in modo che i due potenziali corrispondenti appartengano ciascuno a un gruppo diverso, perchè altrimenti sarebbe possibile annullare un'altra variabile, eliminandola dalla base, senza fare assumere un valore positivo a nessuna delle altre variabili non di base, il che potrebbe avvenire solo se, contrariamente all'ipotesi, non avessimo ancora raggiunto una soluzione-base. Occorre perciò includere in entrambi i gruppi i potenziali corrispondenti ad almeno una variabile non di base.

Il procedimento di cui abbiamo ora delineato i principi fon-

damentali, può essere concretizzato e visualizzato come segue : si parte da una riga analoga a quella dello schema precedente, scegliendo possibilmente una che contenga un segno in corrispondenza di tutti i potenziali. Si dividono poi i potenziali in due gruppi, a seconda dei segni corrispondenti ; chiameremo « potenziali di riga », per esempio, quelli col segno $+$ e « potenziali di colonna » quelli col segno $-$. I potenziali corrispondenti alla variabile non di base il cui simbolo si trova sulla stessa riga, che sono entrambi di colonna, vanno anche inclusi fra i potenziali di riga, in conformità con quanto si è detto sopra. Determinati così i due gruppi di potenziali, si segnano le variabili non di base, in una tabella a doppia entrata, col criterio di simbolizzare la variabile δ_{ij} , nella casella (i, j) se u_i è un potenziale di riga e u_j un potenziale di colonna e nella casella (j, i) nel caso contrario. La variabile alla quale corrisponde la coppia di potenziali ripetuti va segnata sia in riga, sia in colonna. In ognuna delle caselle così determinate, si pone poi uno di quattro simboli diversi, che per esempio possono essere i seguenti: $\boxed{\cdot\cdot}$ $\boxed{\quad}$ $\boxed{\text{I}}$ $\boxed{\text{=}}$ e che hanno rispettivamente questi significati :

$\boxed{\cdot\cdot}$ = il potenziale di riga e il potenziale di colonna si neutralizzano a vicenda (cioè ad ogni aumento dell'uno corrisponde una diminuzione dell'altro, e viceversa, a meno che non si tratti di portare nella base la variabile corrispondente alla stessa casella).

$\boxed{\cdot}$ = entrambi i potenziali sono « neutralizzati da altri potenziali o dai simboli $\boxed{\text{=}}$ e $\boxed{\text{I}}$ rispettivamente

$\boxed{\text{I}}$ = il potenziale riga è « neutralizzato » da un altro potenziale, non così il potenziale colonna ;

$\boxed{\text{=}}$ = il potenziale colonna è « neutralizzato » da un altro potenziale, non così il potenziale riga.

Dal significato dei simboli, risulta evidente che in ogni riga deve figurare, una sola volta, o il simbolo : $\boxed{\cdot\cdot}$ o il simbolo $\boxed{\text{=}}$ i quali si escludono a vicenda. Del pari in ogni colonna deve figurare il simbolo $\boxed{\cdot\cdot}$ o il simbolo $\boxed{\text{I}}$

La rappresentazione di una soluzione-base data in termini di tali simboli, non è unica. Conviene però scegliere una rappresentazione che contenga il numero più elevato possibile di simboli $\boxed{\cdot\cdot}$ nella quale, cioè risulti neutralizzato il maggior numero di po-

tenziali. Inoltre è conveniente collocare lo stesso simbolo nelle caselle corrispondenti alle variabili segnate due volte, che sarà opportuno contraddistinguere, inoltre per esempio con un cerchietto, come in seguito faremo.

Si può quindi procedere nel modo seguente, per ottenere una rappresentazione conveniente: anzitutto si « neutralizzano » coi simboli $\boxed{\cdot}$ i potenziali che figurano sia in riga, sia in colonna; il primo potenziale di riga non neutralizzato si neutralizza con lo stesso simbolo col primo potenziale di colonna, non ancora neutralizzato, a esso collegato dall'esistenza di una corrispondente variabile non di base e così via, fino all'esaurimento delle possibilità di siffatte neutralizzazioni reciproche. Si pone poi il segno $\boxed{-}$ in una casella di ogni riga corrispondente ad un potenziale di riga non neutralizzato, analogamente, per ogni potenziale di colonna non neutralizzato si pone in una casella della colonna corrispondente il simbolo $\boxed{\perp}$. Nelle caselle rimanenti si pone infine il simbolo $\boxed{\cdot}$ (Naturalmente ci riferiamo qui sempre soltanto alle caselle, previamente determinate e corrispondenti a variabili non di base, mentre le altre caselle rimangono vuote).

Si collegano, poi mediante segmenti di retta, tutte le caselle comunque contrassegnate.

La lettura di tale grafico avviene come segue: se si vuole esaminare se è possibile ottenere una soluzione migliore, portando nella base, poniamo le variabili $\delta_{h,k}$ si segue idealmente la spezzata tracciata, a partire dalla casella corrispondente a tale variabile, se in tale casella si incontrano un segmento orizzontale e un segmento verticale. Tale percorso immaginario seguita fino a quando si è esaurita tutta la spezzata, oppure fino a quando si giunge ad una estremità di un segmento in cui la spezzata non formi angolo. Vanno osservate le seguenti regole:

- a) ogni segmento iniziato deve essere percorso interamente;
- b) se si giunge ad una delle caselle corrispondenti ad una variabile rappresentata due volte (cioè in una casella contrassegnata col segno $\boxed{\cdot}$), il percorso deve toccare anche l'altra casella corrispondente alla stessa variabile;
- c) talvolta è possibile compiere due percorsi diversi, a seconda se si comincia dal segmento orizzontale o da quello verti-

cale, uno dei due percorsi essendo però sempre incluso nell'altro. In tal caso ci si limita a considerare il percorso minore.

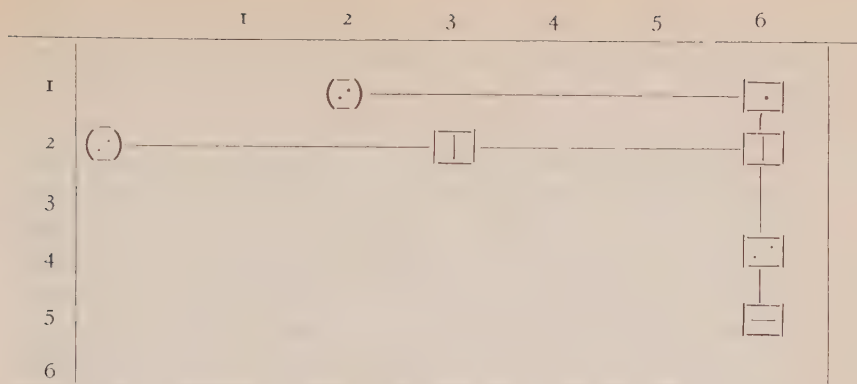
Stabilito il percorso, ci si limita a considerare il numero delle caselle toccate che sono contrassegnate con i simboli $\boxed{\uparrow}$ e $\boxed{\downarrow}$. La caselle contrassegnate coi simboli $\boxed{\uparrow}$ vengono contate col segno $+$ se sono state raggiunte con un percorso diretto iniziato con un segmento orizzontale, oppure se sono state raggiunte col « salto » di cui alla lettera *b*), dopo un percorso iniziato con un segmento verticale. Negli altri casi, invece, vengono contate col segno $-$. L'opposto avviene per le caselle recanti il simbolo $\boxed{\downarrow}$. In altri termini, in ogni percorso diretto, il simbolo corrispondente alla direzione del segmento iniziale assume valore negativo (come deve essere evidentemente, poichè l'introduzione di una variabile nella base comporta la diminuzione di almeno uno dei due potenziali corrispondenti, e di quelli che variano nello stesso senso), quello corrispondente all'altra direzione, invece, assume valore positivo; per le caselle toccate in virtù del salto determinato dal simbolo $(:)$, invece, i segni si invertono. Alla somma algebrica che si ricava dai segni così determinati, se si pone $+1$ per ogni segno $+$ e -1 per ogni segno $-$, si aggiunge poi, se la casella di partenza reca il simbolo $(:)$ o $\boxed{\cdot}$, :

-2 se nel percorso sono comprese sia la riga che la colonna che in quella casella si incontrano;

-1 se è compresa solo la riga o solo la colonna;

Non si aggiunge nulla, invece, per i percorsi che partono da caselle recanti altri simboli, ma entra nel conto naturalmente, con segno negativo anche il simbolo $\boxed{\uparrow}$ o $\boxed{\downarrow}$ che si trovi nella casella iniziale stessa. Se la somma risultante è positiva, si ottiene una soluzione-base migliore introducendo nella base la variabile in questione; se la somma è nulla, tale modificazione sarebbe indifferente, se la somma è negativa, essa sarebbe dannosa. Si verifica immediatamente che le modificazioni utili vanno ricercate soltanto fra quelle che comportano l'aumento di almeno uno dei potenziali corrispondenti ai simboli $\boxed{\uparrow}$ e $\boxed{\downarrow}$.

Nel modo sopra descritto, si ottiene, per esempio, il seguente schema utilizzando la riga relativa alla variabile $\delta_{1,2}$ nella tabella n. 28 e neutralizzando a vicenda i potenziali 1,2 e 4,6.



Dallo schema sopra riportato è facile ricavare tutti i risultati esposti nella tabella n. 28. Ad esempio, per indagare gli effetti di un aumento della variabile non di base $\delta_{2,6}$ si parte dalla corrispondente casella, o lungo il segmento che attraversa la riga 2 o lungo il segmento che si trova nella colonna 6. Nel primo caso si attribuisce il valore $+1$ al segno $\boxed{\cdot}$ posto nella casella $(2,3)$; giunti alla casella $(2,1)$, si passa alla casella $(1,2)$ e si seguono poi i segmenti che si trovano nella riga 1 e nella colonna 6. A causa del passaggio dalla casella $(2,1)$ alla casella $(1,2)$, si attribuisce il valore $+1$ anche al simbolo $\boxed{—}$ posto nella casella $(5,6)$. Il risultato finale è quindi dato dal valore $+2$, che indica la convenienza di introdurre nella base $\delta_{2,6}$. Dallo stesso procedimento si desume anche che in tale modificazione aumenteranno u_3 e u_5 . Agli stessi risultati si giunge, naturalmente, partendo da' segmento contenuto nella colonna 6. Similmente si ottengono tutti gli altri risultati.

Va osservato che l'utilità del metodo che ora siamo andati delineando, appare trascurabile nel caso di problemi di piccola mole, come quello di cui ci siamo serviti a titolo di esempio, mentre diventa abbastanza tangibile in problemi riguardanti un numero più elevato di punti.

Per inciso può essere interessante rilevare che il procedimento esposto prescinde, in certo qual modo, dal valore delle $\delta_{i,j}$ e quindi delle $d_{i,j}$. Ciò sta a dimostrare un fatto abbastanza rimarchevole, e cioè che nei problemi di programmazione lineare, che stiamo trattando, certe soluzioni-base non possono essere mai ottime, quale che sia il valore dei termini noti delle equazioni vincolanti.

Senza tornare sulla questione delle eliminazioni deducibili da una soluzione-base ottima, questione che è già stata trattata nel primo paragrafo, terminiamo ora la parte del presente lavoro dedicata alla esposizione di nuovi procedimenti per la soluzione del problema del commesso viaggiatore, con l'osservare che la conoscenza del percorso ottimo, o, comunque, di un percorso buono, può essere sfruttata in modo ben più efficace di quello esposto nel paragrafo 1, che poi è quello tradizionalmente impiegato. Modificando leggermente il procedimento ora seguito, possiamo infatti giungere a risultati senz'altro più importanti, ammettendo per le variabili corrispondenti al percorso dato, e solo per queste, anche valori negativi. Si può procedere così: nel problema di programmazione lineare, che già avevamo trattato:

rendere massima la funzione $\sum_{i=1}^n u_i$ vincolata alle condizioni:

$$u_i + u_j \leq d_{i,j} \quad (i, j = 1, 2, \dots, n)$$

non si tiene conto delle condizioni il cui termine noto corrisponde a tratti compresi nel percorso dato. Per alcuni di tali tratti, potrà pertanto risultare, se nella tabella finale si tiene conto di tutte le distanze:

$$\delta_{i,j} = d_{i,j} - (u_i + u_j) < 0,$$

Torniamo al nostro esempio, supponendo di avere già individuato il percorso ottimo, dato, come si ricorderà da (1,2,3,4,5,6,

TABELLA N. 29

i/j	1	2	3	4	5	6	u_i
1	—	—	2,6	2	1	—	0
2	—	—	—	2,6	2	3,7	0
3	2,6	—	—	—	2	3,7	0
4	2	2,6	—	—	—	2	0
5	1	2	2	—	—	—	0
6	—	3,7	3,7	2	—	—	0
Min.	1	2	2	2	1	2	

(1)). All'uopo, impieghiamo un procedimento che costituisce una variante, spesso efficace; dell'ultimo procedimento esposto per la determinazione dei valori potenziali. Questa variante si basa sul principio di ridurre, subito, in qualunque stadio un potenziale sulla riga (colonna) del quale venga ad essere iscritto più di uno zero, posti in colonne (righe) non contenenti altri zeri. Contemporaneamente vengono aumentati in eguale quantità i potenziali corrispondenti a tali colonne (righe), fino all'annullamento di un'altra variabile.

Nel nostro esempio, tale procedimento può essere schematizzato come nelle tabelle n. 29 - n. 35.

TABELLA N. 30

i/j	1	2	3	4	5	6	u_i
1	—	—	2,6	2	1,6	—	0
2	—	—	—	0	0	1,1	2,6
3	2,6	—	—	—	2,6	3,7	0
4	2	0	—	—	—	2	0
5	1,6	0	2,6	—	—	—	-0,6
6	—	1,1	3,7	2	—	—	0
Min.	1,6	0	2,6	0	0	1,1	

TABELLA N. 31

i/j	1	2	3	4	5	6	u_i
1	—	—	2,6	0,4	0	—	0
2	—	—	—	0	0	2,7	1
3	2,6	—	—	—	1	3,7	0
4	0,4	0	—	—	—	0,4	1,6
5	0	0	1	—	—	—	1
6	—	2,7	3,7	0,4	—	—	0
Min.	0	0	1	0	0	0,4	

TABELLA N. 32

i/j	1	2	3	4	5	6	u_i
1	—	—	0	0,4	1,6	—	0
2	—	—	—	0	1,6	2,7	1
3	0	—	—	—	0	1,1	2,6
4	0,4	0	—	—	—	0,4	1,6
5	1,6	1,6	0	—	—	—	-0,6
6	—	2,7	1,1	0,4	—	—	0
Min.	0	0	0	0	0	0,7	

TABELLA N. 33

i/j	1	2	3	4	5	6	u_i
1	—	—	0	1,1	1,6	—	0
2	—	—	—	0,7	1,6	1,6	1
3	0	—	—	—	0	0	2,6
4	1,1	0,7	—	—	—	0	0,9
5	1,6	1,6	0	—	—	—	-0,6
6	—	1,6	0	0	—	—	1,1
min.	0	0,7	0	0	0	0	

La soluzione cui si giunge con la tabella n. 36 è una soluzione-base « degenerata » poichè il numero delle variabili nulle è maggiore di n ($= 6$). Di conseguenza dobbiamo considerare una

TABELLA N. 34

i/j	1	2	3	4	5	6	u_i
1	—	—	0	0,3	0	—	0,8
2	—	—	—	0,7	0,8	1,6	1
3	0	—	—	—	0	0,8	1,8
4	0,3	0,7	—	—	—	0	0,9
5	0	0,8	0	—	—	—	0,2
6	—	1,6	0,8	0	—	—	1,1
min.	0	0,7	0	0		0	0

TABELLA N. 35

i/j	1	2	3	4	5	6	u_i
1	—	—	0	0,4	0	—	0,8
2	—	—	—	0	0	0	1,8
3	0	—	—	—	0	0,8	1,8
4	0,4	0	—	—	—	0,1	0,8
5	0	0	0	—	—	—	0,2
6	—	0,8	0,8	0,1	—	—	1,1
Min.	0	0	0	0	0	0,1	

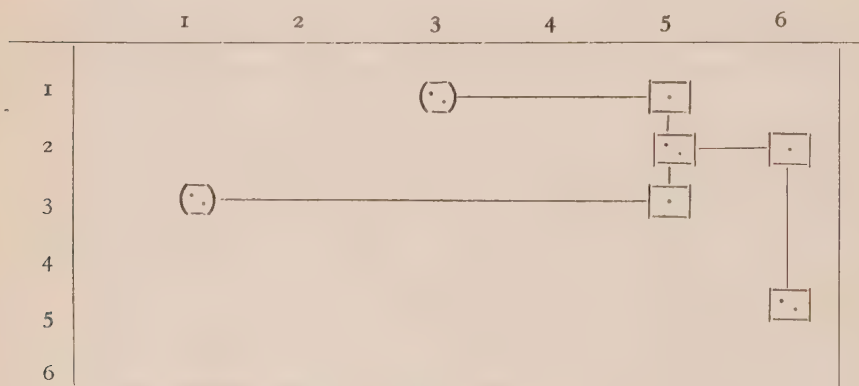
di tali variabili, per esempio $\delta_{3,6}$ come variabile-base nulla. Procedendo allora nel modo esposto precedentemente, e fondandosi,

TABELLA N. 36

i/j	1	2	3	4	5	6	u_i
1	—	—	0	1,1	0	—	0,8
2	—	—	—	0,7	0	0	1,8
3	0	—	—	—	0	0	1,8
4	1,1	0,7	—	—	—	0	0,1
5	0	0	0	—	—	—	0,2
6	—	0	0	0	—	—	1,9

Min. 0 0 0 0 0 0

sempre a titolo di esempio, sulle modificazioni richieste per portare nella base $\delta_{1,3}$ si costruisce lo schema seguente:



In tale schema mancano del tutto i simboli $\begin{smallmatrix} \square \\ \square \end{smallmatrix}$ o $\begin{smallmatrix} \square \\ \square \end{smallmatrix}$ che soli, possono dare un contributo positivo alla somma degli incrementi e decrementi che servono ad indicare la possibilità, o meno, di un miglioramento della soluzione data. Tutte quelle somme, pertanto, sono, nel caso presente, negative o nulle e si può quindi concludere, senza bisogno di ulteriori verifiche, che la soluzione-base cui lo schema si riferisce, cioè la soluzione riportata nella tabella n. 36, è ottima. Non solo, lo stesso schema ci dice anche,

senza altri calcoli, che è possibile giungere ad altre soluzioni-base ugualmente ottime, facendo assumere valori positivi a $\delta_{1,5}$, $\delta_{2,6}$ o $\delta_{3,5}$ in quanto nelle caselle corrispondenti si trova il simbolo « neutro » $\begin{bmatrix} \cdot \\ \cdot \end{bmatrix}$, e quindi la somma degli incrementi e decrementi dei potenziali, relativa a ciascuna di tali variabili, data sempre la mancanza dei simboli $\begin{bmatrix} - \\ - \end{bmatrix}$ e $\begin{bmatrix} \cdot \\ \cdot \end{bmatrix}$, è nulla. Introducendo nella base tali variabili, si ottengono rispettivamente le tabelle n. 37, n. 38 e n. 39.

Se nelle tabelle N. 36 - N. 39 inseriamo anche i valori delle i, j corrispondenti ai tratti che compongono il percorso prescelto, otteniamo rispettivamente le tabelle n. 36a; 37a; 38a; e 39a.

La lunghezza del percorso prescelto (1,2,3,4,5,6,1) espressa in « distanze modificate », $\delta_{i,j}$ è naturalmente sempre uguale,

TABELLA N. 37

i/j	1	2	3	4	5	6	u_i
1	—	—	0	1,1	0,7	—	0,45
2	—	—	—	0	0	0	2,15
3	0	—	—	—	0	0	2,15
4	1,1	0	—	—	—	0	0,45
5	0,7	0	0	—	—	—	0,15
6	—	0	0	0	—	—	1,55

TABELLA N. 38

i/j	1	2	3	4	5	6	u_i
1	—	—	0	0,4	0	—	0,8
2	—	—	—	0	0	0,7	1,8
3	0	—	—	—	0	0,7	1,8
4	0,4	0	—	—	—	0	0,8
5	0	0	0	—	—	—	0,2
6	—	0,7	0,7	0	—	—	1,2

TABELLA N. 39

i/j	1	2	3	4	5	6	u_i
1	—	—	0	0,4	0	—	1,15
2	—	—	—	0	0	0	2,15
3	0	—	—	—	0,7	0,7	1,45
4	0,4	0	—	—	—	0	0,45
5	0	0	0,7	—	—	—	-0,15
6	—	0	0,7	0	—	—	1,55

(e pari a $-3,3$) poichè in tutte le soluzioni-base sopra riportare la somma dei potenziali è la stessa. Variano, invece, i valori delle singole variabili $\delta_{i,j}$ e varia anche, il che, come vedremo subito, è la cosa più importante, la somma rispettivamente dei termini positivi e di quelli negativi del percorso. Tali somme sono date rispettivamente da $+0,8$ e $-4,1$ nella tabella N. 36a; $+1,2$ e $-4,5$ nella tabella n. 37a; $+0,5$ e $-3,8$ nella tabella n. 38a e $+1,3$ e $-4,6$ nella tabella N. 39a. L'importanza di tali somme parziali risulta chiara dalla seguente considerazione: il percorso

TABELLA N. 36a

i/j	1	2	3	4	5	6	u_i
1	—	-0,6	0	1,1	0	-0,7	0,8
2	-0,6	—	-2,6	0,7	0	0	1,8
3	0	-2,6	—	0,1	0	0	1,8
4	1,1	0,7	0,1	—	0,7	0	0,1
5	0	0	0	0,7	—	-0,2	0,2
6	-0,7	0	0	0	-0,2	—	1,9

TABELLA N. 37a

i/j	1	2	3	4	5	6	u_i
1	—	—0,6	0	1,1	0,7	0	0,45
2	—0,6	—	—3,3	0	0	0	2,15
3	0	—3,3	—	—0,6	0	0	2,15
4	1,1	0	—0,6	—	0,7	0	0,45
5	0,7	0	0	0,7	—	0,5	—0,15
6	0	0	0	0	0,5	—	1,55

TABELLA N. 38a

i/j	1	2	3	4	5	6	u_i
1	—	—0,6	0	0,4	0	0	0,8
2	—0,6	—	—2,6	0	0	0,7	1,8
3	0	—2,6	—	—0,6	0	0,7	1,8
4	0,4	0	—0,6	—	0	0	0,8
5	0	0	0	0	—	0,5	0,2
6	0	0,7	0,7	0	0,5	—	1,2

TABELLA N. 39a

i/j	1	2	3	4	5	6	u_i
1	—	—1,3	0	0,4	0	—0,7	1,15
2	—1,3	—	—2,6	0	0	0	2,15
3	0	—2,6	—	0,1	0,7	0,7	1,45
4	0,4	0	0,1	—	0,7	0	0,45
5	0	0	0,7	0,7	—	0,5	—0,15
6	—0,7	0	0,7	0	0,5	—	1,55

prescelto contiene tutti i termini negativi di una tabella, che consideriamo fissata, quindi un tratto al quale, per ipotesi, corrispondesse nella stessa tabella una $\delta_{i,j}$ maggiore della parte positiva del percorso in questione, renderebbe, anche nell'ipotesi migliore, i percorsi in cui figurasse, più lunghi di quello prescelto. Pertanto, tali tratti possono essere senz'altro eliminati. Con lo stesso ragionamento possono essere eliminate anche sequenze di due o più tratti consecutivi.

Ciò fa comprendere l'opportunità di prendere in considerazione soprattutto quelle tabelle in cui la parte positiva del percorso prescelto sia il più possibile ridotta. A tal fine è talvolta opportuna eseguire una combinazione lineare fra due o più soluzioni-base ottime, in modo, appunto, da ridurre il valore di tale parte positiva.

Nel nostro caso, ciò non è necessario; basta considerare la tabella n. 38a, in cui la parte positiva del percorso prescelto è uguale a 0,5. Conviene però, in generale, per individuare rapidamente le eliminazioni possibili, modificare alquanto i valori potenziali, riducendoli, senza però aumentare la parte positiva suddetta. A tale scopo si considera la prima delle righe (colonne) aventi due termini negativi e si aggiunge al potenziale corrispondente il maggiore di tali termini, si opera poi allo stesso modo nella successiva riga (colonna) aventi due termini negativi, tenendo naturalmente conto delle modificazioni già apportate e così via. Alla fine nessuna riga (colonna) dovrà comprendere più di un termine negativo. Nel nostro caso, tale procedimento dà luogo alla tabella n. 38b.

Dalla tabella n. 38b, risulta l'eliminabilità di tutti i tratti corrispondenti agli elementi della seconda e terza riga (colonna) ad eccezione di quelli appartenenti al percorso prescelto, cioè 1-2, 2-3, 3-4. Tali tratti, sono quindi compresi nel percorso ottimo del quale, di conseguenza deve far parte anche il tratto 5-6, ecc. ecc.

Il procedimento esposto è stato qui riportato per la sua vasta applicabilità e efficacia. Nel nostro esempio specifico, tuttavia, si sarebbe potuto giungere allo stesso risultato anche per via diretta, basando il ragionamento sulla tabella N. 38a.

La «lunghezza complessiva» del percorso prescelto, in fatti, in base a tale tabella, è, come abbiamo visto, uguale a —3,3.

TABELLA N. 38b

i/j	1	2	3	4	5	6	u_i
1	—	0	0,6	0,4	0	0	0,8
2	0	—	-1,4	0,6	0,6	1,3	1,2
3	0,6	-1,4	—	0	0,6	1,3	1,2
4	0,4	0,6	0	—	0	0	0,8
5	0	0,6	0,6	0	—	0,5	0,2
6	0	1,3	1,3	0	0,5	—	1,2

Ora, tre soli elementi della tabella sono negativi: $\delta_{1,2} = -0,6$; $\delta_{2,3} = -2,6$; $\delta_{3,4} = -0,6$.

È evidente che un percorso nel quale non figurasse uno solo tratti corrispondenti a tali variabili, avrebbe una « lunghezza » non inferiore a $-2,6 + (-0,6) = -3,2$, cioè sarebbe più lungo del percorso prescelto. Di conseguenza, tutti e tre questi tratti devono fare parte del percorso ottimo, ecc. ecc., In base ad un ragionamento analogo, è, del resto, facile dimostrare, con ognuna delle tabelle N. 36a-39a, almeno l'appartenenza del tratto 2-3 al percorso ottimo.

I due ultimi procedimenti presuppongono, come si è visto, la conoscenza di un percorso, se non ottimo, poco più lungo del percorso ottimo. Si è già accennato nel paragrafo 1 che vi sono metodi per individuare percorsi siffatti. Il metodo più semplice e rapido è probabilmente quello di prendere un punto, unirlo a quello più vicino, e così via. Tuttavia, soprattutto se il numero dei punti non è molto piccolo, l'efficacia di tale metodo appare alquanto dubbia.

Risultati senz'altro migliori, però a costo di calcoli più lunghi possono invece essere ottenuti nel modo seguente: si stabilisce una qualunque permutazione dei punti dati e si determina poi il percorso fra i primi quattro punti che per ognuno di tali punti soddisfa al requisito del « prolungamento » minimo. Poi si prende il quinto punto e lo si inserisce in quel tratto rispetto al quale, il « prolungamento » è minimo. Se il percorso così stabilito soddisfa

anche per tutti gli altri punti presi in considerazione al requisito suddetto, si inserisce il sesto punto ; altrimenti si operano prima quei cambiamenti che sono necessari perchè il percorso dei cinque punti soddisfi al requisito.

In tal modo, si costruisce in definitiva una successione di percorsi, quasi certamente di lunghezza poco maggiore del minimo, per 4,5,6... , n punti. In ultimo, si verifica poi che il percorso di n punti, così determinato, non possa essere migliorato mediante « inversioni », altrimenti si operano tali inversioni e si verifica poi ancora se il requisito dei « prolungamenti » minimi è soddisfatto.

In tal modo, si ottengono in generale dei percorsi che ai fini pratici possono ritenersi quasi equivalenti a quelli ottimi.

Se, anzi si ripete tale procedimento alcune volte, cambiando ogni volta la permutazione dei punti, appare probabile che si possa individuare il percorso ottimo.

Nel nostro esempio, supponiamo di operare con la permutazione 2,5,1,3,4,6. Otteniamo allora successivamente i seguenti percorsi : (1,2,3,5,(1)) ; (1,2,3,4,5,(1)) ; (1,2,3,4,5,6,(1)).

I procedimenti esposti, come è già stato accennato, non devono essere considerati come alternative che si escludono a vicenda, ma, soprattutto quando si tratti di risolvere problemi di una certa mole, come mezzi che possono integrarsi reciprocamente. Ciò può condurre anche a qualche modificazione di qualcuno dei metodi proposti ; così, per esempio, l'impiego della programmazione lineare subisce qualche leggera variazione, peraltro facilmente intuibile, se con altri metodi, oppure anche con lo stesso, si è riusciti a stabilire che certi tratti sono compresi nel percorso ottimo. In tal caso, mentre si trascurano del tutto i punti « interni », di tali tratti, ai potenziali corrispondenti ai punti estremi dei tratti medesimi, sarà conveniente attribuire, nella funzione da rendere massima, un peso pari alla metà di quello che compete agli altri potenziali. L'individuazione di tali tratti dovrà essere del pari tenuta presente nell'impiego di altri metodi di eliminazione e anche in quello tendente a determinare semplicemente un percorso buono. Non si tratta comunque di modificazioni radicali, ma di questioni di dettaglio che un operatore attento non dovrebbe avere difficoltà da risolvere di volta in volta.

TABELLA N. 40

$i \backslash j$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
1	—	49	21	48	26	50	43	41	15	15	97	28	45	54	45	34	1
2	49	—	70	97	74	98	92	28	59	34	53	48	26	103	94	83	2
3	21	70	—	28	11	29	22	61	24	36	118	40	66	34	27	17	3
4	48	97	28	—	39	39	27	88	52	63	145	68	93	41	25	21	4
5	26	74	11	39	—	27	33	53	16	40	109	29	57	43	38	27	5
6	50	98	29	39	27	—	15	75	42	64	131	52	80	17	49	39	6
7	43	92	22	27	33	15	—	83	46	58	140	61	88	14	37	28	7
8	41	28	61	88	53	75	83	—	38	25	57	23	5	92	86	74	8
9	15	59	24	52	16	42	46	38	—	25	94	18	43	58	50	39	9
10	15	34	36	63	40	64	58	25	25	—	82	23	30	69	60	49	10
11	97	53	118	145	109	131	140	57	94	82	—	80	52	148	142	131	11
12	28	48	40	68	29	52	61	23	18	23	80	—	28	68	67	56	12
13	45	26	66	93	57	80	88	5	43	30	52	28	—	96	90	79	13
14	54	103	34	41	43	17	14	92	58	69	148	68	96	—	51	43	14
15	45	94	27	25	38	49	37	86	50	60	142	67	90	51	—	13	15
16	34	83	17	21	27	39	28	74	39	49	131	56	79	43	13	—	16
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	

TABELLA N. 41

<i>ij</i>	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	—	×	—	×	×	×	×	×	—	+	×	×	×	×	×	—
2	×	—	×	×	×	×	×	×	×	+	+	×	×	×	×	×
3	—	×	—	—	—	—	—	×	—	×	×	×	×	×	×	—
4	×	×	—	—	×	×	—	×	×	×	×	×	×	×	+	×
5	×	×	—	×	—	—	×	×	×	×	×	×	×	×	×	—
6	×	×	—	×	—	—	×	×	×	×	×	×	×	+	×	—
7	×	×	—	—	×	×	—	×	×	×	×	×	×	×	×	×
8	×	×	×	×	×	×	×	—	×	×	×	+	+	×	×	×
9	—	×	—	×	—	×	×	×	—	×	×	—	×	×	×	—
10	+	+	×	×	×	×	×	×	×	—	×	×	×	×	×	×
11	×	+	×	×	×	×	×	×	×	×	—	×	+	×	×	×
12	×	×	×	×	—	—	×	+	—	×	×	—	×	×	×	×
13	×	×	×	×	×	×	×	+	×	×	+	×	—	×	×	×
14	×	×	×	×	×	+	+	×	×	×	×	×	×	—	×	×
15	×	×	×	+	×	×	×	×	×	×	×	×	×	×	—	×
16	—	×	—	×	3	—	×	×	—	×	+	×	×	×	+	—
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

3) Alcuni metodi esposti nel paragrafo precedente, sono stati applicati per risolvere il problema di trovare il percorso stradale più breve che unisca 16 capoluoghi di province italiane. I capoluoghi sono, in ordine alfabetico, i seguenti: Ancona, Bari, Bologna, Bolzano, Firenze, Genova, Milano, Napoli, Perugia, Pescara, Reggio Calabria, Roma, Salerno, Torino, Udine, Venezia.

Tali città saranno designate nel seguito per semplicità di scrittura, col numero di ordine che a ciascuna compete nel precedente elenco; così si è posto: Ancona = 1, Bari = 2, Udine = 15, Venezia = 16.

Le distanze stradali, espresse in decine di Km., sono riportate nella seguente tabella N. 40 (1).

Osserviamo anzitutto, che, come accade spesso in simili casi, soprattutto a causa degli arrotondamenti, la disuguaglianza triangolare non è sempre rispettata; così, si ha per esempio: $d_{1,8} = 41 > d_{1,10} + d_{8,10} = 40$. Per quanto riguarda l'eliminazione dei tratti che, unendo due punti, passano per un altro punto, ci siamo attenuti al criterio di eliminare il tratto $P_h - P_k$ quando e, solo quando per un j qualsiasi risulta: $d_{h,k} > d_{h,j} + d_{j,k}$.

Ciò premesso, riportiamo nella tabella N. 41 le eliminazioni e le sub-eliminazioni stabilite col semplice confronto fra gli elementi corrispondenti delle coppie di righe della tabella delle distanze. Il simbolo + posto nelle caselle (h,k) sta ad indicare che il percorso ottimo comprende il tratto $P_h - P_k$.

Tutte le eliminazioni sopra riportate derivano, come è facile verificare, dall'applicazione delle regole esposte nel paragrafo precedente.

Dalla tabella si rileva che già in tale fase sono stabiliti i tratti 1-10-2-11-13-8-12; 4-15-16; 6-14 — 7

Conviene ora considerare i « prolungamenti », soprattutto relativamente ai punti estremi di tratti stabiliti, in quanto, in generale, le eliminazioni fondate sui prolungamenti rispetto a coppie di punti sono più abbondanti che non quelle rispetto a punti singoli. A titolo di esempio riportiamo i calcoli per i « prolungamenti » rispetto alla coppia di punti $P_1 - P_{12}$.

(1) Tali distanze rimontano a qualche anno fa e può darsi che con l'ampliamento della rete stradale, oggi il percorso ottimo risulti non solo più breve, ma addirittura diverso da quello trovato in base ai dati riportati.

TABELLA preparatoria

j	3	5	6	9	16	i/j	3	5	6	9	16
d_{1j}	21	26	50	15	34	3	—				×
$d_{12,j}$	40	29	52	18	56	5	×	—		×	×
						6	×	×	—	×	×
$d_{1,j} - d_{12,j}$	-19	-3	-2	-3	-22	9	×			—	×
graduatoria	4	2	1	3	5	16					—

PROLUNGAMENTI

i/j	3	5	6	9	16
3	—	39	44	15	46
5	39	—	51	28	36
6	44	51	—	25	47
9	15	28	25	—	13
16	46	36	47	13	—

Dai « prolungamenti » si rileva anzitutto che il tr. atto 9-16 può essere eliminato, in quanto corrisponde al prolungamento minimo (e, quindi, risulterebbe più breve il percorso in cui, al posto del tratto 9-16 figurassero i tratti 9-12 e 1-16). Vediamo inoltre che nella stessa tabella il « valore critico » è pari a 25, prolungamento corrispondente alla casella (6,9). Di conseguenza viene eliminato il tratto 5-12. E poichè tutti gli elementi della riga (colonna) 5, sono maggiori di 25, risulta, inoltre, che il punto 9 deve essere unito, nel percorso ottimo, al punto 1 o al punto 12.

Nello stesso modo, mediante la determinazione dei prolungamenti nei rispetti della coppia 4-16, che costituisce gli estremi del tratto 4-15-16, si eliminano i tratti 6-16 e 3-4. In conseguenza di tale ultima eliminazione, 4 deve essere unito a 7 nel percorso

TABELLA N. 42

<i>i/j</i>	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	—	x	—	x	x	x	x	x	—	+	x	x	x	x	x	x
2	x	—	x	x	x	x	x	x	x	+	+	x	x	x	x	x
3	—	x	—	x	—	x	x	x	x	x	x	x	x	x	+	—
4	x	x	x	—	x	x	+	x	x	x	x	x	x	x	+	x
5	x	x	—	x	—	x	x	x	—	x	x	x	x	x	x	x
6	x	x	x	x	—	—	x	x	x	x	x	—	x	+	x	x
7	x	x	x	+	x	x	—	x	x	x	x	+	x	+	x	x
8	x	x	x	x	x	x	x	—	x	x	x	+	+	x	x	x
9	—	x	x	x	—	x	x	x	—	x	x	—	x	x	x	x
10	+	+	x	x	x	x	x	x	x	—	x	x	x	x	x	x
11	x	+	x	x	x	x	x	x	x	x	—	x	+	x	x	x
12	x	x	x	x	x	—	x	+	—	x	x	—	x	x	x	x
13	x	x	x	x	x	x	x	+	x	x	+	—	x	x	x	x
14	x	x	x	x	x	+	+	x	x	x	x	x	—	x	x	x
15	x	x	x	+	x	x	x	x	x	x	x	x	x	—	+	—
16	x	x	+	x	x	x	x	x	x	x	x	x	x	+	—	—
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

ottimo. Poichè già si sa che il punto 7 è unito anche al punto 14, ciò comporta in pari tempo all'eliminazione del tratto 3-7.

I « prolungamenti rispetto alla coppia 6-16 (punti estremi del tratto 6-14-7-4-15-16), portano all'eliminazione dei tratti 3-6, 1-16 e 5-16. Rimane quindi stabilito « il tratto 3-16.

I « prolungamenti » rispetto alla coppia 13-6, punti estremi del tratto 6-14-7-4-15-16-3, conducono alla eliminazione del tratto 3-9. Riassumendo, la tabella delle eliminazioni si presenta a questo punto come segue:

Dalla tabella precedente si rileva anzitutto che nel percorso ottimo deve figurare il tratto 5-9 (in quanto, se il punto 9 fosse unito ai punti 1 e 12, si avrebbe un sub-percorso chiuso). A questo punto si potrebbe arrivare a conclusione, continuando come prima, calcolando cioè i « prolungamenti » rispetto a tale coppia di punti. Giunti in una fase così avanzata, possiamo invece procedere più speditamente, considerando i percorsi che si possono costruire coi tratti non eliminati. È facile rendersi conto che tali percorsi sono soltanto due, e precisamente :

$$a) \ 1(-10-2-11-13-8-) \ 12-6 \ (-14-7-4-15-16-) \ 3-5 \ (-) \ 9- \ (1)$$

$$b) \ 1 \ (-10-2-11-13-8-) \ 12-9 \ (-) \ 5-6 \ (-14-7-4-15-16-) \ 3- \ (1)$$

dove si sono posti in parentisi i tratti comuni ai due percorsi, dei quali non occorre tener conto per la comparazione. Per i tratti che diversificano l'un percorso dall'altro, si ha :

$$a) \ d_{6,12} + d_{3,5} + d_{1,9} = 52 + 11 + 15 = 78$$

$$b) \ d_{9,12} + d_{5,6} + d_{1,3} = 18 + 27 + 21 = 66$$

Risulta quindi che il percorso *b)* è il percorso ottimo. Esso tocca le città nell'ordine seguente : Ancona, Bologna, Venezia, Udine, Bolzano, Milano, Torino, Genova, Firenze, Perugia, Roma, Napoli, Salerno, Reggio Calabria, Bari, Pescara, Ancona (Vedi Figura). La sua lunghezza complessiva è pari a 367, (cioè 3.670 km.)



TABELLA N. 43

i/j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	u_i
1	—	24,5	11	34	6	37	22	21	0	—5	46,5	6	24,5	31	15	8	1
2	24,5	—	54,5	77,5	56,5	79,5	56,5	2,5	38,5	8,5	—3	20,5	0	74,5	58,5	51,5	2
3	11	54,5	—	23	0	25	10	50	18	25	96,5	27	54,5	20	6	0	3
4	34	77,5	23	—	24	31	11	73	42	48	99,5	51	77,5	23	0	0	4
5	6	56,5	0	24	—	13	11	32	0	19	57,5	6	35,5	19	7	0	5
6	37	79,5	25	31	13	—	0	61	33	50	86,5	36	65,5	0	25	19	6
7	22	56,5	10	11	11	0	—	1	29	36	87,5	37	65,5	—11	5	0	7
8	21	2,5	50	73	32	61	61	—	22	4	5,5	0	16,5	68	55	47	8
9	0	38,5	18	42	0	33	29	22	—	9	47,5	0	26,5	39	24	17	9
10	—5	8,5	25	48	19	50	36	4	9	—	30,5	0	8,5	45	29	22	10
11	46,5	—3	96,5	99,5	57,5	86,5	87,5	5,5	47,5	30,5	—	26,5	0	93,5	80,5	73,5	11
12	6	20,5	27	51	6	36	37	0	0	0	26,5	—	4,5	42	34	27	12
13	24,5	0	54,5	77,5	35,5	65,5	65,5	—16,5	26,5	8,5	0	4,5	—	71,5	58,5	51,5	13
14	31	74,5	20	23	19	0	—11	68	39	45	93,5	42	71,5	—	17	13	14
15	15	58,5	6	0	7	25	5	55	24	29	80,5	34	58,5	17	—	24	15
16	8	51,5	0	0	0	19	0	47	17	22	73,5	27	51,5	13	—24	—	16
u_i	1 9,5	2 15	3 0,5	4 4,5	5 10,5	6 3,5	7 11,5	8 10,5	9 5,5	10 10,5	11 41	12 12,5	13 11	14 13,5	15 20,5	16 16,5	

Siamo quindi giunti alla soluzione ottima, senza impiegare i metodi della programmazione lineare. A puro titolo esemplificativo riportiamo comunque, nella tabella n. 43 un sistema di distanze ridotte $\delta_{i,j}$, calcolato secondo il metodo di ammettere valori negativi solo per i tratti di un percorso determinato, che nel nostro caso identifichiamo con quello ottimo.

La « parte positiva » del percorso è pari a 43,5 e quindi si possono eliminare subito 35 tratti, ai quali corrispondono, nella stessa tabella, quantità maggiori. Altre eliminazioni si possono ottenere facendo variare i potenziali u_i , in modo però, da non accrescere la parte positiva suddetta. Così, per esempio, si può eliminare il tratto 12-14, ponendo $u_8 = 10,5 + 2,5 = 13$; $u_{10} = 10,5 + 2,5 = 13$; $u_{12} = 12,5 - 2,5 = 10$; il che comporta:

$$\delta_{12,14} = d_{12,14} - u_{12} - u_{14} = 44,5 > 43,5$$

Risulta tuttavia anche dal presente esempio che tale metodo, in generale, non è sufficiente per risolvere da solo problemi di una certa mole, ma va integrato con altri procedimenti, fra cui possono meritare considerazione quelli illustrati.

BIBLIOGRAFIA

- G.B. DANTZIG, D.R. FULKERSON, S.M. JOHNSON: *Solution of a Large-Scale Traveling-salesman Problem* « Operations Research », Vol. 2° - 1954
- G. MORTON: *A Contribution to the Traveling-Salesman Problem*. « Journal of the Royal Statistical Society ». Vol. 17- n. 2, 1955.
- M. FLOOD: *The Traveling-Salesman Problem*. « Operations Research for Management ». Vol. 2°, Londra, New York
- G.A. CROES: *A Method for Solving Traveling-Salesman Problem*. « Operations Research » Vol. 6°, 1958.
- G.B. DANTZIG, D.R. FULKERSON, S.M. JOHNSON: *On a Linear-Programming Combinatorial Approach to the Traveling-Salesman Problem*. « Operations Research » Vol. 7° - 1959.

CORRADO GINI

**Statistical Methods with special referenc
to Agriculture**

CHAPTER IV
AVERAGES

IV. 1. PREMISES FOR THE MEASUREMENT OF STATISTICAL DATA

IV. 1.1. Data which can be utilised directly.

The data obtained by a statistical survey may be — for the phenomenon to be studied — perfectly accurate and complete, and leave no doubt as to their interpretation. They can therefore be utilised directly. Sometimes, however, this may not be the case, either because there are errors or gaps, or because the data are too rough or because the survey is not complete and the data are not representative of the phenomena or because the data cannot be compared with each other or with other ones.

In the first case, it is possible to proceed at once to measure those aspects of the statistical phenomena which are of interest to us, and for the measurement of which statistical methodology provides us with suitable means. In the second case, it is necessary, in the first instance, to remedy the defects and the imperfections of the data by means of suitable procedures and devices. We shall deal with this second object of statistical methodology in later chapters (chap. XIII and foll.). Here we deal with the measurement of statistical phenomena on the hypothesis that the material at our disposal is accurate, complete and of sure interpretation.

IV. 1.2. Aspects from which a statistical phenomenon can be studied.

The principal aspects from which statistical phenomena can be studied, may be grouped as follows :

1) Determination of the intensity of a phenomenon ; for example, enumeration of the number of inhabitants of a State,

or the number — total or average — of persons employed in a group of agricultural enterprises, or the total amount of the agricultural production of Italy, or its average amount per hectare.

2) Determination of the ratios between several phenomena ; for example, a comparison of the agricultural population or production of Italy with that of Germany, France etc.

3) Description and measurement of the distribution of a collective phenomenon, either with reference to the intrinsic qualities of its component parts (e.g. the distribution of a population classified by its civil status) or with reference to the distribution of its modalities in time (e.g. the population of a State in different years) or with reference to their distribution in space (e.g. the geographical distribution of the population of a State).

4) The description and measurement of the similarity or dissimilarity or other relationships between the distributions of two or more collective phenomena (e.g. if and to what extent the distribution of agricultural enterprises, classified according to the number of persons employed, is similar in Italy to that in one or more other countries).

5) The description and measurement of the relationships between the different modalities of two or more collective phenomena (e.g. given the distribution of a population by a combination of age and occupation to find out if a relationship exists between age and occupations of individuals, and if so, what the relationship is and its degree). Such relationships give rise to multidimensional distributions.

6) The description and measurement of the similarity or dissimilarity of two or more multidimensional distributions (e.g. the similarity or dissimilarity of the distributions of population by a combination of age and occupation in different countries).

IV. 2. STEADY MEANS

IV. 2.1. *Total and mean intensity.*

The intensity of a collective phenomenon may be considered from two points of view, that of total intensity and that of mean intensity. If, for example, we wish to study the production in Italy of a certain agricultural product, for instance wheat,

in a given year, we may want to determine, according to the purpose of our investigation, either its total amount or the average amount per hectare sown to wheat.

The following are some examples of investigations of total intensity, the determination of which, in theory, is very simple.

Total intensities (1952)

Continent	Total Area	Agricultural Area	Total Population	Agricultural Population
	1,000,000 hectares		Millions	
Europe (excluding U.R.S.S.)	493	248	399	128
North and Central America.	2.424	640	201	65
South America	1.778	374	116	68
Asia (excluding URSS).	2.773	782	1 305	913
Africa	3 027	867	209	153
Oceania.	880	397	14	9

To determine mean intensity, which we are now going to deal with, special analytical procedures are required ; they lead to the determination of averages.

IV. 2.2. Concept of average or mean.

Given many quantities, they are either all equal to one another, or at least some are unequal to each other. In the first case the average is represented by any one of them. In the second case there is at least one of them which has the maximum value, and at least one other which has the minimum value, and the average or mean of such quantities either is one of the quantities considered, which is neither higher nor lower than all the other quantities, or a new intermediate quantity among these. In the first case we speak of an *effective mean* or *real mean*, in the second of a *computed mean*.

Another fundamental distinction is that between A) *Steady means* and B) *Unsteady means*, according to whether their value does or does not depend on all the quantities considered. Steady means can be divided in their turn into a) *Basic means*, b) *Ex-*

ponential means and c) *Basic-exponential means*, according to whether in the analytical formulas which express them, the quantities of the group appear as the bases, or as the exponents, or as both bases and exponents.

Each of these averages or means may be simple or weighted. We will explain later the concept of weighted mean.

The above definition gives the *general concept of average or mean*. On the basis of this definition, for every series where the terms differ from one another, an infinite number of averages exist. Thus, in order to identify a certain average, we must specify the operations used to compute one among the infinite number of mean values that are possible. Thus we obtain the *operational concept of average or mean*.

We shall speak further on of the principal operational averages, giving the relative formula for each of them.

IV. 2.3. *The steady basic simple means most widely used.*

The steady means most widely used in Statistics are the basic ones. Among them let us mention the following:

a) *Arithmetic mean*. It is the best known mean; so much so, that, when the word "mean" is used without any other qualification, the arithmetic mean is understood. This mean is given by the sum of the values of the quantities divided by their number. If we indicate the values of the n quantities by a_1, a_2, \dots, a_n , the arithmetic mean will be:

$$M_1 = \frac{a_1 + a_2 + \dots + a_n}{n} = \frac{\sum_{i=1}^n a_i}{n}$$

For example, the arithmetic mean of the following figures giving the production of grapes in Italy in millions of cwts. in the 8 years from 1947 to 1954, and already given in chapter I on page 342, table C.

53, 58, 59, 61, 77, 74, 83, 80

will be:

$$M_1 = \frac{53 + 58 + 59 + 61 + 77 + 74 + 83 + 80}{8} = 68.13$$

b) *Geometric mean*. It is given by the n^{th} root of the product of the values of n quantities. The geometric mean of the values of the quantities considered above will be :

$$M_0 = \sqrt[n]{a_1, a_2 \dots a_n}$$

or more briefly :

$$M_0 = \sqrt[n]{\prod_{i=1}^n a_i}$$

where P is the symbol of product.

In practice the geometric mean is calculated by means of logarithms.

Since

$$\log M_0 = \log \sqrt[n]{a_1 a_2 \dots a_n} = \frac{1}{n} [\log a_1 + \log a_2 + \dots + \log a_n]$$

that is

$$M_0 = \text{antilog} \frac{\sum_{i=1}^n \log a_i}{n}$$

Thus the geometric mean of the grapes produced in Italy in the years mentioned on vol. XX, page 342 will be obtained as follows :

log 53 = 1.72428	
log 58 = 1.76343	
log 59 = 1.77085	
log 61 = 1.78533	
log 77 = 1.88649	
log 74 = 1.86923	
log 83 = 1.91908	
log 80 = 1.90309	
Total	14.62178

$M_0 = \text{antilog} \frac{14.62178}{8} = 67.25$

c) *Harmonic mean*. It is given by the reciprocal of the arithmetic mean of the reciprocals of the quantities. If, as usual, we indicate the generical quantity by a_i , the harmonic mean is given by :

$$\frac{n}{\sum_{i=1}^n \frac{1}{a_i}}$$

Thus the harmonic mean of the figures on page 342 will be given by:

$$M_{-1} = \frac{8}{\frac{1}{53} + \frac{1}{58} + \frac{1}{59} + \frac{1}{61} + \frac{1}{77} + \frac{1}{74} + \frac{1}{83} + \frac{1}{80}} = 66.39$$

d) *Anti-harmonic mean*. It is given by the sum of the squares of the quantities divided by the sum of the quantities; i.e. in symbol by:

$$T_2 = \frac{\sum_{i=1}^n a_i^2}{\sum_{i=1}^n a_i}$$

The formulae of the geometric mean and of the anti-harmonic mean present some drawback when the terms of the series are not all positive. For example, if one term of the series is 0, the geometric mean, being always 0, loses its significance; if the series comprises an uneven number of negative quantities and the total number of quantities is even, the geometric mean loses its significance in the field of real numbers; if the series comprises positive and negative values, the anti-harmonic mean may give values which are outside the range of variation of the quantities, as in the following case:

$$a = -3, a_2 = 2; T_2 = \frac{(-3)^2 + 2^2}{-3 + 2} = -13.$$

As difficulties of this kind occur in many formulae of averages, it must be kept in mind that, where the quantities of the series are not all positive, the formulae should not be used without previous careful examination.

IV. 2.4. *Comprehensive formulae of the steady basic simple means.*

a) *Power means*. A comprehensive formula which gives rise to an infinite number of steady basic simple means is the following:

$$M_r = \sqrt[r]{\frac{a_1^r + a_2^r + \dots + a_n^r}{n}} = \sqrt[r]{\frac{\sum_{i=1}^n a_i^r}{n}}$$

It will readily be understood that from such a formula as many means may be derived as there are values that r can assume, that is an infinite number. For some values of r we go back to some of the means we already know.

One can see at once that, for $r = 1$, we obtain the arithmetic mean and, for $r = -1$, the harmonic mean. For $r = 0$ the preceding expression loses its significance, but when r tends to 0 its limit tends to the geometric mean. For $r = 2$ we obtain the mean known as the *Square mean*, which is widely used in Statistics, especially in calculating errors of observation. It is represented by the expression :

$$M_2 = \sqrt{\frac{a_1^2 + a_2^2 + \dots + a_n^2}{n}}$$

For example, the square mean of the distribution of the production of grapes, some other means of which we have already determined, will be

$$M_2 = \sqrt{\frac{53^2 + 58^2 + 59^2 + 61^2 + 77^2 + 74^2 + 83^2 + 80^2}{8}} = 69.70$$

For $r = 3$ we obtain the *Cubic mean*; for $r = 4$ the *Biquadratic mean* etc.

b) *Mean of power sums*. Another comprehensive formula of steady basic simple means is given by the expression :

$$T_m = \frac{\sum_{i=1}^{i=n} a_i^m}{\sum_{i=1}^{i=n} a_i^{m-1}}$$

known as the *Mean of power sums*. In this formula as well, if we take m equal to 1, we obtain the arithmetic mean. For $m = 0$ we obtain the harmonic mean, and for $m = 2$ the anti-harmonic mean.

c) *Biplane combination power mean*. If, in the formulae of means given above, we substitute for each quantity a_i the product of c ($\leq n$) quantities, and we extend the sum to all

the $\binom{n}{c}$ products corresponding to the combinations of n quantities, taken c at a time, we obtain formulas of *combination means*.

A comprehensive formula of such means is the following which gives the *biplane combination power mean*:

$$B_{c \ p \ q}^d = \sqrt[c p - d q]{\frac{\binom{n}{d} \sum_{i=1}^{\binom{n}{c}} p_i^c (a_i^p)}{\binom{n}{c} \sum_{i=1}^{\binom{n}{d}} p_i^d (a_i^q)}}$$

in which $p_i^c (a_i^p)$ indicates the generic product of the p^{th} power of c quantities and the sum is extended to all the $\binom{n}{c}$ possible combinations out of the n quantities taken c at a time and $p_i^d (a_i^q)$ has a like significance, q indicating the power and d the number of quantities in each of the $\binom{n}{d}$ possible combinations.

For instance the biplane combination power mean of the following quantities,

$$1, 2, 10, 20, 100$$

when $p = 2, q = 1, c = 3, d = 4$, is given by

$$B_{3 \ 2 \ 1}^4 = \sqrt[4]{\frac{\binom{5}{4} \sum_{i=1}^{\binom{5}{3}} p_i^3 (a_i^2)}{\binom{5}{3} \sum_{i=1}^{\binom{5}{4}} p_i^4 (a_i)}}$$

$$\begin{aligned} & \sqrt[4]{\frac{(1 \cdot 2 \cdot 10)^2 + (1 \cdot 2 \cdot 20)^2 + (1 \cdot 2 \cdot 100)^2 + (1 \cdot 10 \cdot 20)^2 + (1 \cdot 10 \cdot 100)^2 + (1 \cdot 20 \cdot 100)^2 + (2 \cdot 10 \cdot 20)^2 + (2 \cdot 10 \cdot 100)^2 + (2 \cdot 20 \cdot 100)^2 + (10 \cdot 20 \cdot 100)^2}{2 \cdot 1 \cdot 2 \cdot 10 \cdot 20 + 1 \cdot 2 \cdot 10 \cdot 100 + 2 \cdot 10 \cdot 20 \cdot 100 + 1 \cdot 2 \cdot 20 \cdot 100 + 1 \cdot 10 \cdot 20 \cdot 100}}} \\ &= \sqrt[4]{\frac{1}{2} \frac{425.241.000}{66.400}} = \sqrt[4]{3202, 11596} = 56,59 \end{aligned}$$

The biplane combination power mean includes as special cases the power mean and the mean of power sums mentioned above, the first of which corresponds to the case where $c = 1, q = 0$, and the second to the case where $c = 1, p = m, d = 1$,

$q = m - 1$. Where $q = 0$, means known as *monoplane combination power means* are obtained, and where $p = q = 1$ the *biplane combination means* are obtained.

It should be noted that the formula denoting the biplane combination power mean may also give, for particular values of the constants and of the quantities, values which are outside the range of variation of the quantities. Such external values, obtained by operating on the quantities in a uniform way, fall within the concept of *numerical adequates* (See Chapter IV.4.4.), while the word "mean" is reserved to those values which do not exceed the maximum and are not exceeded by the minimum quantity of the series.

IV. 2.5. *Weighted means.*

Let us first consider the case in which we have to calculate means of values which occur with different frequencies, giving a frequency seriation.

Distribution of 222 plants of Ranunculus bulbosus according to the number of petals (1).

Number of Petals	Groups of Observations
5	133
6	55
7	23
8	7
9	2
10	2
TOTAL . . .	222

(1) Table taken from VIANELLI, op. cit.

The frequency seriation given above could also be expressed in the following way: 5,5,5, ... (133 times), 6,6,6, ... (55 times) 7,7,7, ... (23 times), 8,8,8, (7 times), 9,9, 10,10, and their sum is evidently given by: $(5 \cdot 133) + (6 \cdot 55) + (7 \cdot 23) + (8 \cdot 7) + (9 \cdot 2) + 10 \cdot 2 = 1250$.

The arithmetic mean will therefore be given by :

$$M_1 = \frac{1250}{222} = 5.63.$$

The frequency of each term expresses its importance in the calculation of the mean. Such frequency is called *weight*, and the mean is known as a *weighted mean* in contrast with the means considered above, which are known as *simple means*. Sometimes even in the case of quantities which do not give frequency seriations, it may be advisable to give the various values different weights. Thus, for example, if we want to estimate the price of next year's wheat and have different forecasts made by different experts, it might be advisable to give the different forecasts different weights, in proportion to the value attached to the reliability of the experts.

If we indicate by a_1, a_2, \dots, a_n the quantities of which we are seeking the mean, and by p_1, p_2, \dots, p_n the corresponding weights, the weighted arithmetic mean will be given by the expression :

$$M_1 = \frac{a_1 p_1 + a_2 p_2 + \dots + a_n p_n}{p_1 + p_2 + \dots + p_n} = \frac{\sum_{i=1}^n a_i p_i}{\sum_{i=1}^n p_i}$$

What has been said for the arithmetic mean holds good for the other means we have spoken of above, and for which we now give the corresponding formulae ;

Weighted geometric mean :

$$M_0 = \sqrt[p_1 + p_2 + \dots + p_n]{a_1^{p_1} \cdot a_2^{p_2} \cdot \dots \cdot a_n^{p_n}} = \sqrt[\sum_{i=1}^n p_i]{\prod_{i=1}^n a_i^{p_i}}$$

Weighted harmonic mean :

$$M_{-1} = \frac{p_1 + p_2 + \dots + p_n}{p_1 \frac{1}{a_1} + p_2 \frac{1}{a_2} + \dots + p_n \frac{1}{a_n}} = \frac{\sum_{i=1}^n p_i}{\sum_{i=1}^n p_i \frac{1}{a_i}}$$

Weighted power mean :

$$M_r = \sqrt[r]{\frac{\sum_{i=1}^n a_i^r p_i}{\sum_{i=1}^n p_i}}$$

It will be readily seen that, if all the weights are positive, as we generally assume, the general expression now given always supplies a value not outside the range of variation of the data, i.e. a true mean. Whenever negative weights are also assumed, a sufficient condition in order to obtain from the formula a true mean value, is that — the quantities being listed in non-increasing order — the following dual limitation be valid for each $s \leq n$:

$$0 \leq \sum_{i=1}^s p_i \leq \sum_{i=1}^n p_i$$

i.e. that the sum of any number s of the weights of the first quantities is not negative and not superior to the total sum of the weights of the n quantities. Said condition however is not necessary as, if it is not satisfied, it is however possible that the expression given above supplies a true mean value.

IV. 2.6. *Exponential and basic-exponential steady means.*

Of less importance than the steady basic means are the exponential and basic-exponential means. The following are their formulae corresponding to the simple arithmetic mean :

$$c^m = \frac{\sum_{i=1}^n c^{a_i}}{n} \quad ; \quad M^m = \frac{\sum_{i=1}^n a_i^{a_i}}{n}$$

(where m is called the *exponential mean with base = c* and M^m is called the *basic-exponential mean*)

and the following corresponding to the weighted mean :

$$c^m = \frac{\sum_{i=1}^m p_i c^{a_i}}{\sum p_i} \quad ; \quad M^m = \frac{\sum_{i=1}^m p_i a_i^{a_i}}{\sum p_i}$$

IV. 2.7. *Dividing value.*

The dividing value of a non-decreasing succession of n quantities a_1, a_2, \dots, a_n is that mean value a_k which gives :

$$\begin{aligned} a_1 + a_2 + \dots + a_{k-1} &< a_k + a_{k+1} + \dots + a_n \\ a_1 + a_2 + \dots + a_k &> a_{k+1} + a_{k+2} + \dots + a_n \end{aligned}$$

If it happens that the following equality remains true :

$$a_1 + a_2 + \dots + a_k = a_{k+1} + \dots + a_n$$

then an infinite number of dividing values are comprised between a_k and a_{k+1} , but in much cases it is conventional usage to take as dividing value the arithmetic mean between a_k and a_{k+1} .

Let us consider, for instance, the following succession :

$$2, 4, 5, 6, 8, 10, 12$$

The dividing value is 8 ; we find, indeed, that

$$\begin{aligned} 2 + 4 + 5 + 6 &< 8 + 10 + 12 \\ 2 + 4 + 5 + 6 + 8 &> 10 + 12 \end{aligned}$$

that is

$$\begin{aligned} 17 &< 30 \\ 25 &> 22 \end{aligned}$$

For the following succession 3, 4, 5, 7, 9, 10, we have instead :

$$3 + 4 + 5 + 7 = 9 + 10$$

The dividing value is therefore any value comprised between 7 and 9. On the basis of what was said above, we shall, however, take as dividing value the arithmetic mean between 7 and 9 i.e. 8.

IV. 2.8. *Note on the calculation of the mean for seriations given by classes of intensity of the character.*

If the character is given by classes of intensity, as in the table shown below, the calculation of the means will be approximate since it is not possible to determine the precise mean value of each class ; this will be effected by taking as intensity of each class the mid-point of the class range, i.a.

3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27. The value of the arithmetic mean, for example, of the given quantities will then be given by :

$$\begin{aligned}
 M &= \frac{1}{1920} (3 \cdot 69 + 5 \cdot 112 + 7 \cdot 149 + 9 \cdot 189 + 11 \cdot 206 + \\
 &\quad + 13 \cdot 241 + 15 \cdot 204 + 17 \cdot 250 + 19 \cdot 224 + \\
 &\quad + 21 \cdot 102 + 23 \cdot 70 + 25 \cdot 53 + 27 \cdot 51) = \\
 &= \frac{1}{1920} (187560 + 1043 + 1701 + 2266 + 3133 + 3069 + 4250 + \\
 &\quad + 4256 + 2142 + 1610 + 1325 + 1377) = 14.026.
 \end{aligned}$$

Group of farms comprised in the Rhine Reclamation Scheme

Size in hectares	Number of farms	Mid-point of class interval	Product
2-4	69	3	187
4-6	112	5	560
6-8	149	7	1043
8-10	189	9	1701
10-12	206	11	2266
12-14	241	13	3133
14-16	204	15	3060
16-18	250	17	4250
18-20	224	19	4256
20-22	102	21	2142
22-24	70	23	1610
24-26	53	25	1325
26-28	51	27	1377
TOTAL .	1.920		M = 14.026

(1) Data taken from VIANELLI, op. cit.

IV. 3. UNSTEADY MEANS.

IV. 3.1. Concept of unsteady mean.

In addition to the various kinds of means of which we have spoken so far and which, as we have seen, may be expressed as functions of all the terms of the series, other mean values exist, the algebraical expression of which does not depend on all the terms of the series ; such values are often represented graphically or by mere verbal definitions. They are the so-called *Unsteady means*, the most important of which will be listed in the following paragraphs.

IV. 3.2. Semi-sum or mid-point of the extreme values.

As may be seen from the example given in the note on § IV. 2.8, this mean is often used in practice, especially when the character is given by classes of intensity.

IV. 3.3. Median.

It is one of the most important mean values. It is that value which bisects the increasing or decreasing scale of quantities in such a way that the quantities preceding it are equal in number to those following it. If the number of quantities is uneven, the median is that quantity, therefore, which occupies the middle place in the scale. If the number of quantities is even, the median is constituted by any value comprised between the two middle quantities of the scale. In such a case it is conventional usage to take the arithmetic mean of the two middle quantities as the median. Thus, for example, the median of the decreasing series 45, 38, 32, 28, 19, is 32; the same value would be the median if the series were an increasing one. In the scale 1, 4, 5, 7, 8, 9, the median is the arithmetic mean between 5 and 7, viz. 6.

Let us now consider the following seriation :

Number of petals	Frequencies	Cumulated frequencies
5	133	133
6	55	188
7	23	211
8	7	218
9	2	220
10	2	222
TOTAL	222	

The median is given by the arithmetic mean between the 111th and the 112th terms. One can see immediately in this case that the 111th and 112th terms pertain to the modality 5 petals which is therefore the median of the given distribution. In the majority of cases, however, in order to determine

the median, it is useful to calculate the cumulated frequencies as indicated in the tables above and below.

If the classification is made on the basis of very large classes, the median value falls in a class that we can call the central class, but it remains for this reason illdefined. Let us take, for example, the distribution on page 346 the arithmetic mean of which ($M = 14.026$) we have already calculated :

Size of farms in hectares	Number of farms	Cumulated frequencies
2-4	69	69
4-6	112	181
6-8	149	330
8-10	189	519
10-12	206	725
12-14	241	966
14-16	204	1170
16-18	250	1420
18-20	224	1644
20-22	102	1746
22-24	70	1816
24-26	53	1869
26-28	51	1920
TOTAL	1.920	

In this case, in order to establish the median value more approximately, recourse is had to hypothesis that the distribution of quantities within the class is uniform, and the following procedure is adopted. By definition, the median value is that value which is exceeded in intensity by 960 cases and whose intensity exceeds that of the other 960 cases, the total number of cases being 1920. The central class is then that of 12 - 14 hectares and, since it has below it 725 cases and above it 954 cases, it follows that on this hypothesis the median value will be given by that one of the 241 cases included in the central class which exceeds in intensity the first $960 - 725 = 235$ cases and is exceeded by the other $960 - 954 = 6$ cases. Therefore the median value is obtained on the hypothesis that the cases are distributed uniformly within the central class, by adding to the value 12 the quantity 1.95 given by the proportion $x : 2 = 235 : 241$ or by subtracting from the value 14 the

quantity 0.05 determined by the proportion $x : 2 = 6 : 241$, where 2 is the interval (14 — 12) of the central class. The result obtained is

$$M = 13.95$$

This procedure for obtaining the median has more or less value according to whether the above-mentioned hypothesis of the uniform distribution of cases within the central class is more or less justified. Such a hypothesis is the more plausible, the smaller is the range of the central class.

IV. 3.4. *Quartiles.*

The *first quartile* (Q) of a seriation, the quantities being arranged in a non-decreasing order is that value below which there is $\frac{1}{4}$ and above which there are $\frac{3}{4}$ of the quantities given; the second quartile (Q) coincides with the median, while the third or upper quartile is that value below which there are $\frac{3}{4}$ and above which there is $\frac{1}{4}$ of the quantities given.

This definition of quartiles, however, in general presents drawbacks since it is often difficult to divide the quantities of the succession in such a way that it is possible to identify that effective or computed term which leaves exactly $\frac{1}{4}$ of the quantities on one side and $\frac{3}{4}$ of the quantities on the other.

It is therefore preferable to define the first quartile as the median of the quantities below the median of the whole seriation, and the third quartile as the median of the quantities above the median.

Thus in the seriation shown on page 347, and of which we have already calculated the median, we find that the first quartile is equal to 5, which is also the median, and the third quartile is equal to 6.

These difficulties however have no practical importance when the number of quantities is not so limited as in the above example.

IV. 3.5. *The mode.*

Another important mean value is the mode. It is given by that value of the character which presents the highest frequency. Thus, for instance, the mode of the distribution on page 346 corresponds to the class 16 - 18.

Finding the mode presents some difficulties when the class intervals are of different size, this because the number of cases in each class depends on the greater or smaller size of the class intervals.

Let us take, as an example, the seriation already given in chapter I, vol. XX page 340, table B.

Size of enterprises	Number of enterprises
0.01- 0.50	909,782
0.51- 1.	581,299
1.01- 3.	1,272,590
3.01- 5.	532,827
5.01- 10.	429,209
10.01- 20.	253,959
20.01- 50.	106,961
50.01-100.	25,575
100.01-500.	17,559
500.01 and above	3,505

It might be thought at first sight that the mode corresponds to the class 1.01 - 3, to which the maximum number (1,272,590) corresponds. But, if we reflect that this class is larger than the preceding ones, we see the need for splitting up the larger classes into smaller ones, so that all the classes have the same amplitude, for example 0.5.

In order to do this, we can have recourse to the hypothesis of a uniform distribution already mentioned on page 100, by means of which we can infer from the data we possess the approximate frequency of class intervals of an amplitude 0.5. The first and the second class in the example do not need to be modified; the third and the fourth must be broken down into 4 classes and so on. It can then be seen at once that the modal value does not really fall in the third class, as it appeared, but in the first.

It should be noted that a distribution may have more than one mode, but that the mean and the median are still unique for multimodal distribution.

Besides the value of highest frequency, or mode, there may be other values which have a frequency higher than the neighbouring values, but lower than the mode. They may be called

relative or *secondary modes*, and in contrast to these, the mode can be called *absolute* or *principal mode*.

IV. 3.6. *The antimode.*

It is given by the value of the character which presents the lowest frequency. A distribution may have more than one antimode. It may also present *relative* or *secondary antimodes* in contrast to the *absolute* or *principal* one.

IV. 3.7. "*Potior*" value.

The "*potior*" value is another important mean value. It is that value of the characteristic which multiplied by its frequency gives a maximum. Let us take, for example, a seriation of families (see table below) classified according to the number of children.

Families resident in the province of Parma classified according to the number of children (1936 census).

Number of children (a)	Number of families (b)	Products a. b.
1	18,235	18,235
2	17,046	34,092
3	11,039	33,117
4	6,516	26,064
5	3,511	17,555
6	1,883	11,298
7	889	6,223
8	361	2,888
9 or more	260	2,340

While the mode of the seriation is 1, since the maximum frequency corresponds to that value, the "*potior*" value is 2. This value, in fact, multiplied by its frequency, gives a product (34,092) which is higher than any of the other products obtained by multiplying the number of children by the respective frequency. In this case the "*potior*" value indicates the class of families which contributes most to the succeeding generation.

IV. 4. FURTHER OBSERVATIONS

IV. 4.1. *Continuous means.*

The definitions already given of means refer to the case in which the quantitative modalities are discontinuous. They are easily extended to the case in which they represent real numbers continuous in the interval a to b , by substituting the corresponding integrals for the sums. Thus, for instance, the simple arithmetic mean becomes in a continuous case :

$$M = \frac{\int_a^b x \, dx}{\int_a^b dx} = \frac{1}{2} (a + b)$$

IV. 4.2. *Typical means*

The concept of typical mean varies from one author to another, since each has his own concept of "type". We shall adhere to the following definition : the typical mean is the mode, whatever be the frequency of the phenomenon on either side of the mode itself.

IV. 4.3. *Objective and subjective means.*

The mean may be calculated for the purpose of obtaining the sole and real intensity or value of the phenomenon which is being studied ; this is evidently the case when there is only a single value of the phenomenon, or, in other words, when there is a single objective magnitude to which the different measurements refer. Thus, when the height of a tower, of a mountain or of a star is observed many times by the same or by different persons, the diversity of the observations is due to accidental causes (state of the atmosphere, unsteadiness of the hand, error in reading the result) and possibly to systematic causes (imperfection of the instruments and personal tendencies of the observers). If we suppose that only accidental causes are operative, or that they alone are of importance, the mean of the measurements obtained can be calculated in order to eliminate their influence and to determine the probable value of

the real magnitude. In such a case, therefore, the mean gives us, or at least it is held to give us, the objective value of the magnitude and for this reason it is called an *objective mean*.

If, on the contrary, different objective magnitudes exist, the calculation of the mean will clearly not have the same purpose, nor will it give the same result. If, for example, we have data concerning the number of employees of a group of agricultural enterprises, the mean of these data is not objective. It may, in fact, happen that there is a firm which has a number of employees equal to the mean of the employees of the group of enterprises observed, but this coincidence will be purely accidental and, indeed, it may not take place at all. Such a mean is called *subjective*.

IV. 4.4. *Numerical adequates.*

Some mathematicians define, as the mean of the quantities a_1, a_2, \dots, a_n in relation to a function f , the quantity M which satisfies the relationship

$$f(M, M, \dots, M) = f(a_1, a_2, \dots, a_n)$$

This definition is wider than the definition we have given, in as much as it includes not only the values which do not exceed the extreme quantities (*internal means*), to which statisticians reserve the name "mean", but also values external to them (called by some authors *external means*). We do not believe this definition is acceptable both because the definition "external mean" (that is external intermediate value) is contradictory, and because external means do not fulfil the purpose of a mean, which is to measure the intensity of a collective phenomenon.

To the quantity M , as defined above, we give the name of *numerical adequate*.

IV. 4.5. *Deviations*

The differences observed between the terms of a series and one of its mean values are called *deviations* from the mean in question, or *errors*.

IV. 4.6. *Properties of means.*

When the intensity of a collective phenomenon has to be measured, any of the various means described above cannot be chosen arbitrarily, since the different means have different properties which, according to the cases and the nature of the research, may make one preferable to another.

We have already spoken of some properties common to categories of means. Particularly important is the property of the steady means of being influenced by all the quantities to which the mean refers and, on the contrary, of the unsteady means of remaining indifferent to the variations of some of them.

As far as individual means are concerned, the arithmetic mean has four very important properties; these are:

a) the algebraical sum of the deviations from the mean is nil;

b) the sum of the squares of the deviations is at a minimum compared with sum of the squares of the deviations from any value other than the arithmetic mean;

c) the arithmetic mean of a seriation a_1, a_2, \dots, a_n with weights p_1, p_2, \dots, p_n is the resultant of a system of forces of intensity equal to the p_i , applied at the points a_i of a horizontal axis;

d) it constitutes the centre of gravity (barycentre) and of suspension of a horizontal axis to which are applied at the points of the axis a_i , forces of gravity of intensity equal to the p_i .

The median possesses the following properties:

a) the number of positive deviations is equal to the number of negative deviations;

b) the sum of the absolute values of deviations from the median is at a minimum compared with the sum of the deviations from any value other than the median;

The mode has the property of maximising the number of nil deviations.

The semi-sum or mid-point of the extremes possesses the property of minimising the maximum deviation.

The geometric mean also possesses important properties:

a) the reciprocal of the geometric mean of various quantities is equal to the geometric mean of the reciprocals of these quantities;

b) the geometric mean of the ratios of corresponding quantities in two series is equal to the ratio of their geometric means ;

c) the m^{th} power of the geometric mean of various quantities is equal to the geometric mean of the m^{th} powers of these quantities ;

d) the logarithms of the geometric mean of n quantities is equal to the arithmetic mean of the logarithms of these quantities.

The last of these properties, as we have already seen, greatly facilitates the calculation of the geometric mean.

The power mean M_r has the following properties :

a) it is an increasing function of the index r , that is to say that

$$\dots M_{-1} \leq M_0 \leq M_1 \leq M_2 \leq M_3 \dots$$

the signs of equality being valid only in the case in which all the terms are equal ;

b) if r tends to $-\infty$ or $+\infty$, it tends to the smallest of the quantities or to the largest of the quantities considered respectively.

The mean of power sums T_m possesses the following property : it is an increasing function of the index m , that is to say

$$T_m \geq T_{m-1}$$

the sign of equality being valid only in the case in which all the quantities are equal.

IV. 4.7. Use of means.

When we were speaking of the various properties of the means we said that, in order to measure the intensity of a collective phenomenon, one cannot choose a mean arbitrarily ; on the contrary, the mean must be chosen, which, because of its peculiar properties and in view of the purpose of the research, seems to be most adapted to the particular case. This is a fundamental principle which must always be born in mind.

If the mean has an objective value, that is to say if it aims at ascertaining the probable value of a magnitude really existing in nature, and, if there is reason to believe that measurements

taken of this magnitude are affected by errors of observation, then the mean to be used will be determined by the law which the deviations follow. If these have an accidental character it is usual to admit that they compensate one other; on such hypotheses it is advisable to adopt the arithmetic mean, since this equalises the positive and negative deviations.

If, on the contrary, the errors are not of an accidental nature, then we shall have to choose the mean in conformity with the law that their distribution follows. The law of the errors, however is very often unknown. Since for many phenomena it is represented by a curve analogous to that to which accidental errors conform (Gaussian curve), the hypothesis that this curve is generally valid is often admitted. In this lies one of the reasons for the widespread use of the arithmetic mean.

Gini has, however, demonstrated that the principle of the compensation of accidental errors is not valid in itself. Errors are called accidental when there is no reason to believe that their causes act in one direction rather than in the other; this justifies us in believing the causes of the accidental errors compensate one another, but not that their effects, i.e. the errors are likewise compensating. Such compensation takes place in reality only on particular hypotheses, which should strictly be studied from case to case to see how far they correspond to reality.

A more general hypothesis would indeed lead to the adoption of the mode; it would be sufficient in this case to admit that the greater were the errors, the rarer would they be.

If, on the contrary, the mean has a subjective value, if, that is to say, it aims at giving a synthetic measurement of the individual values that are comprised in the collective phenomenon under study, since it is known that, whatever mean value is chosen, it must differ more or less from the effective values for the phenomenon, it is preferable, in general to choose that mean value which differs as little as possible from these values.

According to the purpose of the inquiry, it is however necessary to establish in what sense and to what extent it is important that the mean value selected differs as little as possible from the effective values of the phenomenon. It may, indeed, be important that the sum of the squares of the deviations be minimum, or that the algebraical sum of the deviations be nil.

In such cases, the proper mean to select will be the arithmetic mean, since it possesses exactly these properties. It may, on the contrary, be important that the sum of the absolute values of the deviations be minimum and in that case the median should be preferred, since it precisely fulfils that condition.

On other occasions it may be important to select from the the series that individual value which appears as typical, in which case recourse will be had to the mode, and so on.

When all the quantities are equally reliable, a steady mean, which takes account of all of them must be preferred, *coeteris paribus*, to an unsteady one, but it may occur that certain quantities are more likely to be affected by errors than others, and then it may be advisable to have recourse to an unsteady mean, which is not influenced by them. For instance, it is possible that the highest or the lowest values or both be suspected to be partly due to errors of observation or, although corresponding to reality, to be abnormal in character. In these cases, a mean like the mode or the median, which are not influenced by such extreme quantities, may be preferable to the arithmetic mean or other steady mean.

Sometimes it is desirable to use a mean for more than one purpose; for instance, the mean of commodity prices may be used as a direct index of the cost of living or of the general level of prices and as an inverse index of the purchasing power of money, and then it may be advisable to adopt a mean that gives concordant results for the two inverse phenomena, as is the case for the geometric mean and would not be the case for the arithmetic mean.

If the quantities are limited in number and represent a random sample of a more comprehensive mass and the main interest is in the measurement of the characteristics of the mass rather than of the sample, account must be taken of the probable random error to which the mean is exposed, and it may be advisable to choose the mean whose probable error is lower. This condition, however, is not necessarily satisfied by the same mean for different kinds of distributions; for the Gaussian distribution, this criterion leads to the choice of the arithmetic mean. In a following chapter we shall treat of accidental errors of statistical constants due to the limited number of observations.

Generally this formal consideration will be taken into account only subordinately to the substantial considerations developed above ; it may make advisable the calculation at the same time of more than one mean rather than the exclusive calculation of that mean which is affected by the lowest random error.

In conclusion, we shall say that it is impossible to lay down any fixed criterion as to which mean should be preferred ; the choice of the mean depends essentially on the purpose of the inquiry.

It is also to be remarked that in many cases the choice of one rather than another of the usual means (arithmetic, geometric, harmonic, median) does not lead to substantially different results. But if the quantities considered present a very wide range of variation (as for instance the prices of commodities during the strong inflation which in some countries followed the first world war), the results may be quite different according to the kind of mean adopted and in that case it would always be a wise procedure to calculate various kinds of means at the same time.

IV. 4.8. *Mean values in correlated series.*

We now pass to the study of another important question.

There are cases in which we find two or more series of quantities referring to characters which depend on each other, such as those of prices and consumption. These series are commonly said to be linked to each other by a law of correlation. In such cases it must be borne in mind that *the mean values of the various series do not necessarily correspond* ; that is when the mean of a series has been determined on the basis of a particular criterion and the mean of another series is calculated on the basis of the same criterion, the latter does not necessarily correspond to the former.

Given a series of prices and the corresponding consumptions and supposing constant conditions of quality and environment, Gini has found that consumption varies approximately in inverse proportion to the logarithm of the prices ; this means that, if the prices rise or fall according to a geometrical progression, consumption falls or rises respectively according to an arithmetical progression. Let the price series be 1,2,4,8,16, that is to say form

a geometrical progression with a ratio of 2, and the consumption series on the other hand 5,4,3,2,1, that is to say form an arithmetical progression with a difference of 1. In order to find the average consumption we make use of the arithmetic mean and we obtain the value 3. If the average price corresponding to such average consumption were now required, it might be thought that just as the average consumption was obtained by calculating the arithmetic mean of the consumption series, so the price corresponding to the average consumption might be obtained by calculating the arithmetic mean of the prices. This would lead to an erroneous result, that is to a price of 6.2, which does not correspond at all to the average consumption of 3. This happens because, the nature of the relationship existing between prices and consumption is not such as to justify the use for the arithmetic mean for determining the average price corresponding to the average consumption. To determine this the geometric mean must be used. In the above example, the geometric mean gives an average price of 4, which is precisely the third term of the price series, and corresponds to the third term (3) of the consumption series.

What has been said for price and consumption series can be repeated for any other couple of series linked by a correlation law.

In general, it may be said, that if terms of two series are linked by a constant relationship, when a mean value has been assumed for the former, the mean value to be assumed for the latter is not arbitrary, but has to be determined from the nature of the relationship existing between the terms of the two series. It may be added that, if such a relationship is linear, then the arithmetic mean of the first series corresponds to the arithmetic mean of the second. But if, on the contrary, the variations of the one are directly or inversely proportional to the variations of the logarithms of the other, then the geometric mean of the latter corresponds to the arithmetic mean of the former; and, finally, if the variations of the one are proportional to the reciprocal of the variations of the other, the harmonic mean of the latter corresponds to the arithmetic mean of the former.

What has been said about correlated means is very important in connection with the so-called *theory of the average man*. This

theory was upheld by Quetelet, then for some time considered to be without foundation, but Gini has eventually shown it to be well-founded.

The theory states that, if all the characteristics of a population are attributed to one man, with values corresponding to the arithmetic means of such characteristics, one obtains the so-called *average-man* who appears as the type of the population itself. In another of his works, however, Quetelet, studying the relationship existing between the two anthropometric characteristics, weight and stature, made the assertion that weight increased in proportion to the square of the stature. If that had been true, the theory of the average man could not have stood, since it would have been impossible for the means of the two characteristics weight and stature to correspond. Hence it followed that either the theory of the average man was false or the stated relationship between weight and stature was false. In the past students of the question held the theory of the average man to be false and the relationship between weight and square of the stature to be true. However, more extensive studies showed that it was precisely this relationship between weight and stature as propounded by Quetelet that was false, while the theory of the average man could very well be maintained. It was in fact found that, if in a diagram of Cartesian co-ordinates the stature is represented on the x -axis and the average weight of the persons with that stature on the y -ordinate, the resulting curve is approximately a straight line; this means that, taking the average of the cases, we find that the variations in the weight are proportional to the variations in stature and that consequently the arithmetic means of the two characteristics do, in fact, correspond. And since this type of relationship holds not only for the stature and weight, but also for all anthropometric characteristics studied so far, it may well be said that not only is Quetelet's average man admissible, but also that, taking the averages of all the somatic characteristics, they can suitably be used to represent the type of the species. On the basis of measurements of American soldiers which were taken during the first world war, a statue was constructed: this statue has a perfectly normal appearance and affords an adequate picture of the adult American man.

IV. 4.9. *Abbreviations used in calculating the arithmetic mean.*

Special devices exist which give considerable economy in the work of calculating the arithmetic mean. It is for example sufficient to observe that the arithmetic mean is equal to any value Z added to the arithmetic mean of the deviations of the given quantities from the said value Z , viz :

$$M_1 = Z + \frac{\sum_{i=1}^n (a_i - Z)}{n}$$

If Z has a value which is presumed to be near the mean, the calculations will be greatly abbreviated, as the following example shows.

Let us consider once again the distribution of the 1920 farms comprised in the Rhine Reclamation Scheme, the arithmetic mean of which we have already calculated by the direct method (see pages 345-346). Assuming $Z = 15$, the calculations will be arranged as follow :

x_i	f_i	$x_i - 15$	$(x_i - 15) f_i$
3	69	-12	-828
5	112	-10	-1120
7	149	-8	-1192
9	189	-6	-1134
11	206	-4	-824
13	241	-2	-482
15	204	0	0
17	250	2	500
19	224	4	896
21	102	6	612
23	70	8	560
25	53	10	530
27	51	12	612
	1920		-1870

$$M_1 = 15 + \frac{-1870}{1920} = 15 - 0.974 = 14.026$$

Such a method is the more economical, the smaller are the deviations from Z of the actual quantities considered.

Another device can be utilised when the classes, on the basis of which the classification is made, differ among themselves by several units but are all of the same class-interval, as in the following example.

*Distribution of European countries according to the percentage of
agricultural population (1926 - 1940)*

Percentage of agricultural population	Number of countries
0-10	3
10-20	0
20-30	11
30-40	3
40-50	2
50-60	5
60-70	1
70-80	3
80-90	1
TOTAL	29

In this distribution all the classes have the same interval 10.

Classes which differ amongst themselves by one unit should be substituted for the original classes. We thus obtain the following table, from which we find that the arithmetic mean is 3.41. In order to obtain the true mean it is only necessary to multiply 0.41 by the ratio between the real and the auxiliary amplitude of the class (10 : 1) and add the result to the central value of the class corresponding to 3, viz. 35.

Thus we shall have :

$$M_1 = 35 + 0.41 \cdot 10 = 39.1$$

x_i	f_i	Product $x_i f_i$
0	3	0
1	0	0
2	11	22
3	3	9
4	2	8
5	5	25
6	1	6
7	3	21
8	1	8
TOTALS	29	99

$$\frac{99}{29} = 3.41$$

IV.5. EXTENSION OF THE CONCEPT OF AVERAGE TO QUALITATIVE CHARACTERS.

IV.5.1. General introduction.

So far, when speaking of the average we have always referred to quantitative characters. But it is also possible to extend the concept of average to the field of qualitative characters. Let us see how this can be done.

When speaking about the concept of average of a character with reference to quantitative characters we said that by average, in the wide sense, might be understood any value which is not exceeded by, and which does not exceed all the measurements of the characters, that is any value not exceeding the maximum modality and not being exceeded by the minimum modality. This implies that in order to speak of the average of a character we must have a highest modality and a lowest modality of it. Now, in the case of qualitative characters, this condition only holds good for rectilinear series while it does not hold good for cyclical and unconnected series. In the case of rectilinear series, a quantitative value is made to correspond to any modality of the qualitative character, by which means we can then operate, for the purposes of averages, on the transformed character as though it were, in fact, a real quantitative character. In the case of cyclical and unconnected series, this cannot be done and therefore one cannot speak of averages with reference to these in the same way in which one speaks of average with reference to quantitative characters.

Nevertheless we can extend the concept of average also to such kinds of series; this may be done on the basis of a principle widely used in the mathematical sciences and which is called "principle of the conservation of formal laws". The principle is as follows: — "For a class of terms, a certain operation is defined which possesses certain characteristic properties, so that, in the same class of terms, the definition may be substituted by such properties possessed by the operation. If the class of terms is amplified so that a new class is obtained for which the definition of the given operation loses its significance, but for which a new operation can nevertheless be defined, which in the new class possesses the same properties and may thus be identified

with the operation in the original class when it is applied to the new class, then the denomination that is used for the operation in the original class is attributed to the new class”.

The extension of the concept of average to qualitative series is very important; it is, in fact, desirable that one should be able to express the characteristic quality of a collective phenomenon with a single modality; for example, to be able to indicate the average day of the week on which marriages are contracted, the average month in which births take place, the average culture of a territory, the average profession of a population.

In the paragraphs which follow we shall see how the arithmetic mean, the median and the mode of cyclical and unconnected series can be determined on the basis of the principle stated above.

IV. 5.2. Extension of the concept of average to cyclical series.

Deviations in cyclical series. — It is very important to decide how deviations in cyclical series may be determined, as doubts and uncertainties may arise. For example, how can the difference between Saturday and Wednesday be calculated? It can be measured in two ways. One can, in fact, pass from Saturday to Sunday and thence to Monday, Tuesday and Wednesday, in which case the difference is 4 days, or, on the other hand, one can pass from Wednesday to Thursday, to Friday and thence to Saturday, in which case the difference is 3 days. Conventionally, however, distances are measured in the way in which they prove smallest, so that between Wednesday and Saturday the difference would be three days and not four.

At first sight such a procedure may seem to be arbitrary, but in reality it is not so at all. It may, moreover, be said that measuring distances in the way in which they prove smallest not only is a matter of convenience, but corresponds to an instinct of ours. We say, in fact, that certain object, being, as we are, at the surface of the earth, is, for example, 5 metres away from us, while just as exactly we might say that it is 40 million minus 5 metres away, if, rather than measuring the distance by the shortest way we measured it in the way which

proves the longest, in this case calculating all the remaining part of the earth's meridian which separates us from the object.

It is interesting to remark that our eye, when observing a wheel which turns, connects its successive positions according to the contiguity of the successive impressions received even if they are not in the same sense of the positions really occurred. This is the reason why in the cinema the wheels of the cars appear sometimes to us as turning in the inverse sense.

It may be understood, moreover, that, for the calculation of the deviations of the modalities of a cyclical series from one of these taken as centre, it is convenient to consider the difference by which the modality follows the centre as positive, and the distance by which the modality precedes it as negative. For example, let us take Sunday as the modality selected as centre, that is the modality from which we wish to calculate the deviations; the minimum distance between Monday and Sunday, which represents the deviation of Monday from Sunday, will be $+1$, while that of Saturday from Sunday will be -1 .

If the cycle has an even number of modalities, there will be one modality from which the distance from the centre will be equal in absolute value whether one proceeds in a positive or a negative direction; thus the deviation of this modality may be indicated either by the sign $+$ or the sign $-$. In such circumstances one conventionally agrees that half the cases corresponding to that modality be indicated with the sign $+$ and the other half with sign $-$.

Arithmetic mean of a cyclical series. When we were speaking previously of the arithmetic mean of the quantitative characters we said that it possesses the following properties: 1) That the algebraic sum of the deviations is nil; 2) that the sum of the squares of the deviations is a minimum. These properties are characteristic of the arithmetic mean so that it is possible to replace the definition of the arithmetic mean by indicating either one of these properties.

On the basis of the principle of the conservation of formal laws, and bearing in mind the definition already given of deviation, we can say that the arithmetic mean of a cyclical series is that modality of the cycle which possesses one of the two properties set forth above.

It is easy to realise that, while in quantitative series there can be only one arithmetic mean, in qualitative there can be more than one, and if one starts from the first property, there may be none at all.

To obviate the latter difficulty it is usual to have recourse to computed modalities and suppose that the succession of the modalities represent a continuum. This is, on the other hand, analogous to what is usually done in the case of quantitative characters. For example, we have said that the arithmetic mean of the number of petals of *Ranunculus bulbosus* is 5.63 which represents a computed and not an effective modality, since there is obviously no plant which has a fractionary number of petals.

When one starts from the second property, there is always an effective modality for which the sum of the squares of the deviations is a minimum. Nevertheless, for sake of generality, the hypothesis of a continuum is introduced also in this case.

When recourse is had to such an hypothesis, the sum of the squares of deviations from the various modalities can present, as well as one or more absolute minima, several relative minima. It has been shown that among the values which satisfy the first property of the arithmetic mean, only one or some of them correspond to the said absolute minima, the others corresponding to relative minima.

We have already pointed out above that, as well as the two properties already examined, the arithmetic mean also possesses a third property. This is: 3) the arithmetic mean constitute the resultant (or barycentre) of the system of forces applied to the points of a horizontal axis, corresponding to the values of the terms with intensities corresponding to the weights attributed to such terms, and consequently the centre of suspension of said axis.

Giving this significance to the arithmetic mean we can represent the cyclical series in two ways: 1) by means of points on a circumference on which the various modalities are marked, in such a way that at each point corresponding to a modality a weight proportional to its frequency is applied; 2) by means of vectors issuing from the centre of the same circle according to the direction of the points corresponding to the various modalities and of length proportional to their frequency.

In relation to the first of these representations it can be said, owing to the principle of formal laws, that any centre of suspension of the circle represents an arithmetic mean of the given series. By centre of suspension is meant any such point that, if the circle were suspended from that point, it would hang in such a way that the centre of the circle, because of the effect of the weights applied to the circle, would fall on the vertical passing through the said point.

If, on the other hand, the other representation is assumed, the resultant of this system of vectors can be determined and will be found to have a certain length and a certain direction that will cut the said circumference in a certain point. This point may be termed the barycentre in so far as it constitutes the point of application on the circumference of the resultant of the system of vectors which have been assumed to represent the frequencies or other magnitudes corresponding to the modalities of the series.

Each of these two conceptions is adapted to certain kinds of cyclical series more than to others.

When, on the basis of the second mechanical interpretation, the resultant of the system of vectors has been determined by analytical process, it will be found that the point at which the resultant cuts the circumference, is also one of the centres of suspension of the same circumference, the other centre being the point diametrically opposite.

On the basis of this last fact we can conclude that the two said conceptions of the arithmetic mean coincide, in the sense that there is a single point on the circumference which constitutes both the barycentre and also one of the two centres of suspension.

We shall now illustrate the calculation of the arithmetic mean of some cyclical series in some simple cases.

We have a cyclical frequency series, for example the number of persons going to a certain market according to the days of the week, and we establish that the total number of buyers in the week under consideration was equal to 250, of whom 30 on Monday, 50 on Tuesday, 20 on Wednesday, 80 on Thursday, none on Friday, 10 on Saturday and 60 on Sunday. Knowing the reciprocal deviations of the days of the week amongst themselves, as seen from the following table ;

From/to	Mon.	Tues.	Wed.	Thur.	Fri.	Sat.	Sun.
Mon.	0	1	2	3	03	02	01
Tues.	01	0	1	2	3	03	02
Wed.	02	01	0	1	2	3	03
Thur.	03	02	01	0	1	2	3
Fri.	3	03	02	01	0	1	2
Sat.	2	3	03	02	01	0	1
Sun.	1	2	3	03	02	01	0

and multiplying the deviations of each modality (day of the week), for the respective frequencies, we shall find that the algebraic sum of the deviations thus obtained is, taking Monday as origin, equal to $1 \times 50 + 2 \times 20 + 3 \times 80 - 3 \times 0 - 2 \times 10 - 1 \times 60 = 250$. On the other hand it will be equal to zero for Tuesday, 180 for Wednesday, -10 for Thursday, -50 for Friday, -50 for Saturday and -60 for Sunday. The definition by which, even in the case of qualitative characters the term which had the property of making the algebraic sum of the deviations zero would be termed the arithmetic mean of the series, thus holds for the term "Tuesday" and for no other day but "Tuesday". Tuesday can therefore be called the arithmetic mean of the qualitative character "day of the week" in the cyclical frequency series considered.

Let us now consider this other cyclical series :

Number of days of hail according to the months (1)
(Bavaria, 1890 - 1929)

<i>J</i>	<i>F</i>	<i>M</i>	<i>A</i>	<i>M</i>	<i>J</i>	<i>J</i>	<i>A</i>	<i>S</i>	<i>O</i>	<i>N</i>	<i>D</i>
40	48	132	374	636	735	646	439	221	78	50	40

In this case no effective modality satisfies the first property of the arithmetic mean. As a matter of fact the algebraic sums of the deviations present, from the various months, the following values :

(1) As the month reduced to 31 days.

from	January	+ 4280
"	February	+ 7331
"	March	+ 7832
"	April	+ 6167
"	May	+ 3476
"	June	+ 557
"	July	— 2422
"	August	— 5353
"	September	— 7732
"	October	— 8155
"	November	— 5434
"	December	— 547

Let us now calculate the arithmetic mean on the basis of the second property. The squares of the deviations between the terms will appear from the following table :

	J	F	M	A	M	J	J	A	S	O	N	D
J	0	1	4	9	16	25	36	25	16	9	4	1
F	1	0	1	4	9	16	25	36	25	16	9	4
M	4	1	0	1	4	9	16	25	36	25	16	9
A	9	4	1	0	1	4	9	16	25	36	25	16
M	16	9	4	1	0	1	4	9	16	25	36	25
J	25	16	9	4	1	0	1	4	9	16	25	36
J	36	25	16	9	4	1	0	1	4	9	16	25
A	25	36	25	16	9	4	1	0	1	4	9	16
S	16	25	36	25	16	9	4	1	0	1	4	9
O	9	16	25	36	25	16	9	4	1	0	1	4
N	4	9	16	25	36	25	16	9	4	1	0	1
D	1	4	9	16	25	36	25	16	9	4	1	0

and, multiplying the squares of each deviation by the respective frequencies, we obtain the sum of the squares of the deviations from the effective modalities.

From	January	it is	71.522
"	February	" "	58.669
"	March	" "	42.198
"	April	" "	27.341
"	May	" "	17.530
"	June	" "	13.437
"	July	" "	15.502
"	August	" "	23.125
"	September	" "	36.714
"	October	" "	54.053
"	November	" "	69.334
"	December	" "	75.785

so that the arithmetic mean of the cyclical series given, considering only the effective modalities, is the month " June ".

Having now recourse to the hypotheses that the distribution is a continuum we find that the sum of the positive deviations is equal in absolute value to the sum of the negative ones for a computed modality corresponding to about the 7th June.

In correspondence with the same computed modality the sum of the squares of the deviations reaches its minimum ($=13.253$) which, obviously, is something lower than that (13.437) corresponding to an effective modality. Seven of June may then be considered the arithmetic mean of the cyclical series given, both starting from the first as from the second of its proprieties.

Let us now consider a third cyclical series concerning the

Number of days of rain according to the months (1)

Pisa (1878-1932)

J	F	M	A	M	J	J	A	S	O	N	D
552	567	642	645	551	372	221	265	472	689	701	701

Also in this case no effective modality satisfies the first property of the arithmetic mean, the algebraic sums of the deviations being

from January	+	407
" February	—	3085
" March	—	5071
" April	—	4513
" May	—	2401
" June	—	217
" July	+	893
" August	+	1199
" September	+	2045
" October	+	3359
" November	+	4127
" December	+	3257

The second property is satisfied by the effective modality "January" the sums of squares of deviations being

from January	590.15
" February	619.57
" March	713.55
" April	822.41
" May	894.07
" June	918.45
" July	902.75
" August	882.73
" September	854.79
" October	800.93
" November	720.43
" December	635.85

(1) As the month reduced to 31 days.

If we have recourse to the continuum-hypotheses, we find that in this case the first property of the arithmetic mean is satisfied by not less than 7 computed modalities corresponding approximately to the following days: 11 and 22 January, 25 May, 4 July, 5 August, 6 September and 13 October.

All these computed modalities correspond to relative minima of the sum of the square of deviations (1), but that of 22 January (=58.526) represent the absolute minimum satisfying the second property of the arithmetic mean. This is somewhat lower than that (59.015) obtained for the effective modality "January."

Mode and antimode of a cyclical series. The usual definitions of mode and antimode, valid for quantitative series, are also applicable in cyclical series. In a cyclical frequency series the mode will in fact be that modality or one of them, to which the maximum frequency corresponds and the antimode the modality or one of them, to which the minimum frequency corresponds.

Median of a cyclical series. That term of the series in relation to which the sum of the absolute deviation values is at a minimum is defined as the median.

From the series of the number of days of hail in Bavaria given on page 368 it is seen that the sum of the absolute deviations calculated

from January	is	11.186
" February	"	10.705
" March	"	9.508
" April	"	7.683
" May	"	5.760
" June	"	4.941
" July	"	5.452
" August	"	7.127
" September	"	9.128
" October	"	10.359
" November	"	10.758
" December	"	11.163

so that the median is the modality, "June", being equal to the arithmetic mean.

It is not necessary to introduce computed modalities, as the minimum of a sum of the absolute deviations always corresponds to an effective modality.

(1) These minima are 58,883, 58.546, 88.792, 89.362, 87.516, 84.969, 80.062 respectively. One can realize that they represent relative minima comparing those figures with those representing the sums of squares deviations from effective modalities.

For the series of the days of rain in Pisa, given at page 370, the sums of the absolute deviations are the following :

from	January	15.293
"	February	15.441
"	March	15.215
"	April	15.269
"	May	16.285
"	June	17.343
"	July	18.517
"	August	18.015
"	September	16.549
"	October	15.175
"	November	14.471
"	December	14.637

Here the median modality is " December ". It is not identical with the arithmetic mean which, as we have seen, is " January ".

IV. 5.3. Extension of the concept of average to unconnected series.

In view of their nature, it is impossible in the case of unconnected series to say that the difference between two modalities is greater or less than the difference between two others. It is instead necessary to allow that the difference between any two of them is constant, that is the successive modalities whatever be their order, are equally spaced. Such constant differences may be represented by a constant positive number, for example the positive unit, which may be termed "deviation". This established, we see how the concept of arithmetic mean may be extended to unconnected series.

If we limit ourselves to seeking an effective modality in the series, which may fulfil the properties of the arithmetic mean, the possibility of there being one which makes the algebraic sum zero must be excluded, since all the deviations are positive.

As far as the property of minimising the sum of the squares of the deviations is concerned, it is seen that such a property is found in that modality or in any one of them, the frequency of which is at a maximum. This modality, therefore, constitutes the arithmetic mean of the unconnected series in accordance with the second property.

Finally, when we have recourse to the mechanical interpretation, it is clear that there does not exist any effective modality which can be the barycentre of the system of points repre-

senting the various modalities, since for a rational representation it is necessary to represent the n modalities of the unconnected series in the vertices of a regular n -sided figure in a space of $n - 1$ dimensions, giving each vertex a weight equal to the frequency of the corresponding modality.

If, however, we introduce computed modalities, it is possible to find a modality which satisfies the above properties, substituting, where the first property is concerned, the concept of vector for the concept of deviation. The computed modality in such case has for coordinates in the space of $n - 1$ dimensions the frequencies of the n effective modalities.

It is observed that such computed modality is different from the effective modality that satisfies the property of minimizing the sum of the squares of the deviations, and that the sum of the squares of the deviations from such effective modality is superior to the respective sum for the computed modality in question.

The *mode* and *antimode* of an unconnected series are, as usual, the modalities to which the maximum and minimum frequency correspond respectively.

The *median* of an unconnected series is that modality in relation to which the sum of the absolute values of the deviations is at a minimum, that is to say that, as can easily be seen, the median of an unconnected series is that modality to which the maximum frequency correspond. It coincides consequently with the mode and the arithmetic mean. For example, the arithmetic mean, the median and the mode of the distribution given in chapter I, vol. XX page 345, table H are constituted by the modality, "fodder"; and those of table E, chapter I, by the modality, "Roman pecorine cheese".

It is worth while to call the attention of the reader to the fact that, when the means are extended to qualitative characteristics on the basis of their properties, several of them (arithmetic mean, median, mode) which are different for quantitative characters or rectilinear series, coincide (so far as we consider effective modalities) for unconnected series and, on the contrary, the arithmetic mean, which is unique for the quantitative characters or rectilinear series, can split into several different modalities for the cyclical series.

CHAPTER V

STATISTICAL RATIOS

V. 1. DIFFERENCE BETWEEN THE INTENSITIES OF TWO COLLECTIVE PHENOMENA.

V. 1.1. Necessity for homogeneity.

The total or average intensities of two collective phenomena between which there is a logical connection, or of the same phenomenon in two different places or times, can be compared by calculating either the difference or the ratio between the two intensities.

In order that the two quantities may be compared, they must be homogeneous. It would be meaningless, for example, to calculate the difference between two average temperatures, one expressed in Centigrade and the other in Fahrenheit. It is necessary, therefore, that the two intensities be expressed in the same unit of measurement; but this condition is not sufficient to give meaning to any difference, for example, it is difficult to see, the meaning of the difference between the price of corn and the price of electrical energy, both expressed in liras.

V. 1.2. Absolute and relative differences. Statistical ratios.

Differences between two intensities may be divided into absolute and relative differences. Absolute differences express the excess or the deficit of one intensity compared with the other. In general, however, the same absolute difference has different degrees of importance according to the terms to which it refers. For example, a difference of 50.000 liras between the cost of two consignments of goods may be relatively more important than a difference of 300.000 liras if the costs compared are, for example, 10.000 and 60.000 liras in the first case, and 700.000 and 1.000.000 liras in the second case. In every case the comparison between the difference and the terms themselves must be explicit, and be effected by relating the difference either to the minuend or

to the subtrahend or to a mean of these, generally the arithmetic mean.

Indicating by I_1 and I_2 the intensities of two collective phenomena, where $I_1 > I_2$, $D = I_1 - I_2$ is the absolute difference and

$$D' = \frac{I_1 - I_2}{I_1} ; D'' = \frac{I_1 - I_2}{I_2} ; D''' = \frac{I_1 - I_2}{\frac{1}{2}(I_1 + I_2)}$$

are the relative differences.

Relative differences also have the advantage, compared with absolute differences, of not being dependent on the unit of measurement in which the two intensities are expressed: the possibilities of comparison are thereby increased. It would, in fact, be meaningless to compare the absolute differences between the unitary prices of corn at two different times with the differences between the unitary prices of iron at the same times, while it may very well be said that the price of corn has increased by 30 % and that of iron by 40 %, in which case the two (relative) differences are perfectly comparable.

Given $\frac{I_2}{I_1} = R' \frac{I_1}{I_2} = R''$ we obviously have $D' = I - R'$;

$D'' = R'' - I$.

This shows that the relative differences between two quantities may easily be expressed by means of the ratios between these quantities. This justifies the fact that the methodology of the theory of differences has developed so little compared with that of statistical ratios.

V. 2. *Classification of statistical ratios.*

Just as statistical differences must refer to phenomena between which there is a logical connection, so must statistical ratios; it would be meaningless, for example, to divide the quantity of corn imported during a certain period by the number of deaths from cancers. But ratios can be established even between quantities that are not homogeneous, (e.g. between the yield of corn and the area sown with corn during a particular period) or between a collective phenomenon and one that is not collective, (e.g. between population and the area of the State).

V. 2.I. *Classes of ratios.*

Viewed as concepts, ratios are divided into two large categories :

A) *Ratios which give rise to a concept which is analogous to that expressed by one of the two terms* ; for example, the ratio between the number of births and the population gives rise to the concept of birth-rate or natality, which is analogous to that of the first phenomenon ;

B) *Ratios which give rise to a concept different from that expressed by either of the two terms* ; for example, the ratio between the amount of the deposits on hand and the amount paid in or drawn out in a unit of time gives rise to the new concept of average length of deposit.

The procedure by which a statistical ratio is determined, is different according to whether it fall in class A) or class B).

Class A) ratios are said to be *simplified*, when one of the two terms is made equal to a fixed quantity as 1, 10, 100, 1000 ; for example, the ratio between the number of births and the population is simplified by assuming the number of inhabitants equal to 100 or 1000 and calculating the number of births there would be in that case.

These ratios can be simplified in two ways ; what is to say that either of the two terms may be made the fixed term. It is, however, more convenient to adopt as the variable term the phenomenon of which we wish to study the variations, (in our case the births), making the other term equal to 1 or 10 or 100 or 1000.

The ratios in the second class, on the other hand, are said to be *solved*, since one finds the quotient of the two quantities which measure the intensities of the two phenomena compared.

V. 3. RATIOS THAT CAN BE SIMPLIFIED.

V. 3.I. *The arithmetic mean.*

The arithmetic mean is the ratio between the aggregate amount of a phenomenon and the number of cases in which the phenomenon occurs. Such a ratio is simplified by making

the second term equal to unity, so that we obtain the intensity of the character that, given the aggregate amount observed, would be applicable to every statistical unit, if each of these were of identical intensity.

V. 3.2. *Index numbers.*

Index numbers solve the problem of how to compare the intensity of a phenomenon in different times or places, by comparing all its intensities with the same intensity, the latter being usually the intensity of the phenomenon in one of the times or places considered, or an average of these. Having made this intensity (*basic intensity*) equal to a power of 10 (generally 100 or 1000), the index numbers are obtained by dividing all the intensities by said basic intensity and then multiplying the quotients by the selected power of 10. These indices are called *fixed base indices*. According to whether the various intensities of the phenomenon refer to different times or places, we have *time indices* or *geographical or place indices*.

Let us consider as an example prices in francs per quintal of oats in France in the years indicated below. If we assume the price in the year 1938 as base, we obtain the time series of index numbers shown in the third column of the following table :

Price of oats

Year	Price per Quintal in Francs	Index Number
1938	118.43	100
1939	84.27	71.16
1940	145.00	122.44
1941	214.00	180.70
1942	266.00	224.61
1943	268.75	226.93
1944	269.00	227.14
1945	472.00	398.55
1946	701.00	591.91
1947	1383.00	1167.78
1948	1815.00	1532.55
1949	1817.50	1534.66
1950	1625.25	1373.33
1951	2515.83	2124.32
1952	2220.83	1875.23

If, on the contrary, we take the production of potatoes in the various continents in 1952 and make the production in Europe equal to 100, we obtain the place series of index numbers shown in the third column of the following table :

	Production (in millions of metric tons.)	Index number
Europe	126.3	100
North and Central America	11.4	9 03
South America	5.0	3.96
Asia	10.4	8.23
Africa	1 0	0.79
Oceania	0 5	0 40

In general, place indices may be considered to form non-ordered series since there is no objective criterion by which to establish a natural series of all the indices. In time series, however, there is a natural order of succession determined by the time. In this case, instead of all the intensities being compared with the same one, each can be compared with the previous one ; in this way a system of *variable base* or *chain indices* will be obtained instead of a fixed base index.

This system, applied to the prices of oats given in the above table gives rise to the series of chain indices shown below.

Price of oats

Year	Chain Indices
1938	100
1939	71.16
1940	172 07
1941	147 59
1942	124 30
1943	101.03
1944	100 09
1945	175 46
1946	148.52
1947	197 29
1948	131.24
1949	100.14
1950	89 42
1951	154.80
1952	88.27

Such *variable base* or *chain indices* clearly show the intensity of the phenomenon from one year to the next, but do not allow an immediate comparison between the intensities of two non-successive years. In order to compare, for example, the price for the year 1945 with that for the year 1940, when the chain indices have been calculated, it is necessary to multiply together all the chain indices from 1945 to 1941, after each has been divided by 100. The index of 1945 with base 1940 is obtained. In general, if a_1, a_2, \dots, a_n are n successive intensities of the phenomenon, the corresponding chain indices are :

$$i_1 = 100, i_2 = \frac{a_2}{a_1} \cdot 100, i_3 = \frac{a_3}{a_2} \cdot 100 \dots i_n = \frac{a_n}{a_{n-1}} \cdot 100$$

and hence

$$100 \frac{i_{r+1}}{100} \cdot \frac{i_{r+2}}{100} \dots \frac{i_{r+k}}{100} = 100 \frac{a_{r+1}}{a_r} \cdot \frac{a_{r+2}}{a_{r+1}} \dots$$

$$\dots \frac{a_{r+k}}{a_{r+k-1}} = 100 \frac{a_{r+k}}{a_r}$$

is the index of the year $r + k$ with base the year r .

It is not possible to calculate fixed base index numbers when all the data are not such as may be compared amongst themselves. For example, monthly statistics on the employment of workers in industries or in agriculture do not always cover the same number of establishments or firms every month. If for the months of January and February 5000 firms are considered and for March only 4,800 firms, the variation in February relative to January cannot be compared with that in March relative to February since the two variations do not refer to the same number of establishments. It is, therefore, impossible to calculate index numbers for February and March taking January as fixed base. If however, we also know the number of workers for the month of February in the same 4,800 firms covered for the month of March, we can relate the March data to these data for February and compile chain index numbers, that is to say, the February index number with January as base (relating to 5000 firms) and the March index number with February as base (relating to 4,800 firms).

The two index numbers are not comparable with each other in the strictest sense, but if the difference between the number of firms taken into account in the two months is small, as in this case, and the firms taken into account for one index number and not for the other do not differ systematically, one can act on the hypothesis that, in a given period, the relative variation for the firms not considered is equal to that for those taken into account, in which case the two chain indices may be compared with each other.

If, however, the difference between the numbers of firms taken into account in the two index numbers had been large or the firms taken and not taken into account had been systematically different, then the hypothesis would not have been admissible and it would have been impossible to compare the two chain indices with each other.

The index numbers of a phenomenon, for example those of prices of oats quoted above, are also called *simple index numbers* to distinguish them from *complex index numbers*, which are means (generally arithmetic or geometric) of the simple index numbers relative to various phenomena.

Complex index numbers may be either *synthetic* or *composite*, according to whether the simple index numbers which contribute to their formation all relate to phenomena of the same species (for example, prices of more than one kind of goods, production of different articles, consumption of various food-stuffs, etc) or, on the contrary, of different species (for example, marriages, foreign trade, production, consumption etc.) which may have to be considered in order to measure the variations of very complex phenomena, such as, for example, the general economic situation of a State.

The *general index of prices* is a synthetic index which serves to measure the variations in the general level of prices, and hence also in the purchasing power of the money, which is its reciprocal: it can be determined in various ways; for example, it may be expressed as the arithmetic mean of the simple indices for the various kinds of goods, each index being weighted according to the value ($= \text{price} \times \text{quantity}$) of fixed quantities of the respective goods at the base time or place. If the prices at the base time or place 0 are p_{0i} , and those at the

base time or place 1 are p_{1i} , and if we indicate the fixed quantities by q_{ki} , the form of the index is then as follows:

$$i_{1,0} = \frac{\sum_{i=1}^m \frac{p_{1i}}{p_{0i}} p_{0i} q_{ki}}{\sum_{i=1}^m p_{0i} q_{ki}} = \frac{\sum_{i=1}^m p_{1i} q_{ki}}{\sum_{i=1}^m p_{0i} q_{ki}}$$

that is to say, the ratio between the values of the fixed quantities of goods at the prices of time or place 1 and at the prices of base time or place 0. The system of the quantities q_{ki} constitutes the *type* of the index. If the comparison is extended to several places or times, with no change in the system of the quantities we have a system of *fixed indices*. If, on the other hand, the system of the quantities q_{ki} is changed in successive comparisons we have a *sliding type index*.

For the purpose of measurement of variation in prices, the method of weighting indicated is inconvenient in that, whenever there is no type system of quantities q_{ki} of particular significance, as there is in the case of the cost of food, there are two systems of quantities to which one may have recourse, namely q_{0i} and q_{1i} and hence two weighting systems $p_{0i} q_{0i}$ and $p_{0i} q_{1i}$. Since we generally have $q_{1i} \neq q_{0i}$ the two indices which are obtained also prove different. This difficulty is eliminated by constructing the two separate indices which are obtained in this way and then calculating one of their means. Theoretical considerations suggest the geometric mean, to which the term *ideal index numbers* has been applied by Irving Fisher, as shown in the following formula:

$$I_{1,0} = \frac{\sum_{i=1}^m p_{1i} q_{0i}}{\sum_{i=1}^m p_{0i} q_{0i}} \times \frac{\sum_{i=1}^m p_{1i} q_{1i}}{\sum_{i=1}^m p_{0i} q_{1i}}$$

When the comparison is extended to n times or places, this formula can however give rise to serious difficulties as it does not satisfy the time or place *circular test* which asks that the index for a time or place B in relation to A multiplied by the index for a time or place C in relation with B gives the index of C in relation to A. It is therefore advisable to use

Gini's formula obtained from the geometric mean of n indices, in each of which the quantities exchanged in a particular time or place are assumed as weights, so that the index of the place or time r in relation to the place or time s is given, whatever the terms r and s are, by the formula :

$$I_{rs} = \sqrt[n]{\frac{\sum_{i=1}^m p_{ri} q_{1i}}{\sum_{i=1}^m p_{si} q_{1i}} \cdot \frac{\sum_{i=1}^m p_{ri} q_{2i}}{\sum_{i=1}^m p_{si} q_{2i}} \cdots \frac{\sum_{i=1}^m p_{ri} q_{ni}}{\sum_{i=1}^m p_{si} q_{ni}}}$$

We will now apply Gini's formula.

Dr. C. Benedetti has made the following application of the formula for the calculation of the price indices of food-stuffs in 7 Italian towns, viz. Bolzano, Lecce, Salerno, Sassari, Messina, Campobasso and Milan in 1938; the foodstuffs taken into account are those forming the normal daily ration in each town of a worker in the various industries, arts and crafts.

In our case :

— p_{ri} and q_{ri} are respectively the price and the quantity of the foodstuffs i in the town r ;

— n is the number of towns (in our case $n = 7$);

— m is the number of foodstuffs composing the average daily ration (in our case $m = 17$).

The following table gives the values of the ration of each town according to the prices of the 7 towns, that is the value of :

$$\sum_{i=1}^m p_{ri} q_{si} \text{ with } \left. \begin{matrix} r \\ s \end{matrix} \right\} 1^{st}, 2^{nd}, 3^{rd}, 4^{th}, 5^{th}, 6^{th}, 7^{th} \text{ town.}$$

Towns	Bolzano 1)	Lecce 2)	Salerno 3)	Sassari 4)	Messina 5)	Campo- basso 6)	Milan 7)
Bolzano 1)	3.984	4.313	4.087	4.091	4.781	4.084	3.949
Lecce 2)	2.740	2.718	2.930	2.499	2.541	2.955	2.947
Salerno 3)	3.082	2.984	3.057	2.934	3.269	3.073	3.128
Sassari 4)	3.383	3.431	3.489	3.392	3.768	3.390	3.649
Messina 5)	3.570	3.629	3.715	3.188	3.911	3.696	3.878
Campobasso 6)	3.064	3.028	3.193	3.024	3.424	3.115	3.523
Milan 7)	4.559	4.516	4.656	4.457	5.192	5.763	4.738

Each number in the table represents the sum of the products of the prices of the m foodstuffs in the town named at the head of each column by the corresponding quantities of the town named on the mother column.

Thus, if we wish to obtain, for example, the index I_{rs} where r = Messina and s = Milan, we have only to multiply the numbers in column 5) by one another, multiply the numbers in column 7) by one another, divide the first product by the second and then find the root 7^{th} .

That is :

$$I_{Mess.Mil} = \sqrt[7]{\frac{\text{Products of numbers in column 5} = 10404.123019}{\text{Product of numbers in column 7} = 8616.661564}} = \\ = \sqrt[7]{1.10744245} = 1.027$$

Generally, by dividing the product of the numbers in the h^{th} column by that of the k^{th} column and finding the 7^{th} root, we shall obtain I_{hk} .

For the seven towns we shall have as many indices I_{hk} as there are simple combinations of 7 elements 2 by 2, viz $\binom{7}{2} = 21$. If the inverse indices are required (that is I_{hk} instead of I_{kh}) and the unity indices ($I_{hh} = 1$) in order to construct a table, obviously we shall have to consider $7^2 = 49$ indices.

It is necessary to distinguish clearly between the *type* and the *base* of a series of index numbers. In a series of chain index numbers, the base of each index number is constituted by the preceeding index, while the type *may* remain the same. The type may, however, alter, as it would do in the above example in which the firms are 5000 in the February index number with January as base and 4.800 in the March index number with February as base. As we have said, one can pass from this chain index series to a fixed base series and calculate the March index number with January as base. It is easy to change the base of a series of index numbers by means of a simple proportion, but the type cannot be changed without recomputing all the indices.

V. 3.3. Coexistence ratios.

Another category of ratios which may be simplified is that of *coexistence ratios* by means of which the intensities of the same phenomenon in different places, or the intensities of two phenomena in the same place are compared.

If we calculate the ratio between the value (in thousands of liras) of Italian goods exported to Europe and those exported to America (period 1st Jan–27th May 1946) we obtain $\frac{9,783,220}{4,243,704} = 2.31$ which tells us that the importance of Italian exports to other European countries is more than double that to American countries. This is an example of the first of the two above types of coexistence ratios. This type of coexistence ratio may also be considered a place index number.

If, instead, we calculate the ratio between the total value in thousands of liras of Italian imports and that of Italian exports (period 1st Jan–27th May 1946) we obtain $\frac{30,002,838}{14,880,159} = 2.02$ which shows that Italy's imports are roughly double her exports. This is an example of the second of the above two types of coexistence ratios.

Other examples of coexistence ratios are grouped in the following table :

	Arable land (incl. fallow and orchards)	Permanent meadows & pastures	Coexistence ratios
	Millions of (a)	Hectares (b)	(b)/(a)
Europe (excluding U.S.S.R. and Turkey) .	150	98	0.65
North and Central America.	256	384	1.50
South America	67	307	4.58
Asia	349	433	1.24
Africa	244	623	2.55
Oceania.	21	376	17.90

V. 3.4. *Composition ratios.*

Composition ratios compare the intensities of two phenomena one of which can be considered part of the other, or — in other words — the intensity of one modality of the phenomenon and the intensity of the whole phenomenon.

The examples of composition ratios grouped in the following table are obtained from the data on agricultural population given in Chapter I, vol. XX page 342. In the second column are given the ratios between the agricultural population and the total population, and in the third column the ratios between the agricultural area and the total area of the various continents.

	Agricultural Pop./Total Pop.	Agricultural Area/Total Area
Europe (excluding U.S.S.R.)	0.32	0.50
North and Central America	0.29	0.26
South America.	0.59	0.21
Asia.	0.70	0.85
Africa.	0.73	0.29
Oceania	0.29	0.46

V. 3.5. *Derivation ratios.*

Derivation ratios compare the intensity of a phenomenon with that of another which it necessarily implies. For example, the phenomenon of births necessarily implies that of population. If we compare the number of live births in the year 1941 in Italy with the average population in the same year (either the population at the middle of the year or the arithmetic mean of the population at the beginning and at the end of the year may be taken) we obtain the following result, if based on a population of a thousands:

$$\frac{937.546}{45,100,000} \times 1000 = 20.8 \text{ per thousand.}$$

This particular derivation ratio is called the *natality rate* or *birth-rate*, and expresses the number of live births in one year per thousand inhabitants.

The same phenomenon may be considered to imply more than one phenomenon, and vice-versa. For example, births imply population, but in a population there are people who, on account of their age, cannot produce children, so that the ratio might be established between the number of births and the number of inhabitants between the ages of 15 and 70. Since such a process is one of specification, we shall call the first ratio the *general derivation ratio* and the second the *specific derivation ratio*. In both cases, a phenomenon is compared with another which it implies; but, in the first case, it is a question of a general implication, while in the second case of a more specific one. When the specification reaches the limit, beyond which greater specification would not be possible, we have a *causal ratio*.

In the above example, the *derived phenomenon* is dynamic in character, that is an event (birth), while the *implied phenomenon* is a static one, that is a condition (living persons): a derivation ratio of this kind is called a *rate*. Thus the ratio of births, of deaths or of marriages to inhabitants gives the birth-rate, the death-rate or the marriage-rate; the ratio of deaths to sick people gives the lethal rate, the ratio of divorces to married people the divorce-rate.

The number of events obviously increases with the time elapsed, so that every rate must be referred to a certain interval of time. If no time is specified, it is generally understood that the interval is a year.

The event may be unique, as in the case of death, or susceptible to repetition as in the case of marriage.

Usually the number of cases of the derived phenomenon considered as the numerator is inferior to that of the implied phenomenon considered as the denominator, but the contrary case is not excluded when the event is susceptible to repetition, as when we divide the number of head-colds by the number of inhabitants.

The implied phenomenon can, however, also be dynamic in character, and in that case, it is generally an event with which the implied condition begins. So we may compute a derivation ratio between the number of deaths and that of births, births being the event with which the condition of the living person

starts. Similarly, we may compute derivation ratios by dividing the divorces by the marriages, or the deaths by the cases of illness. In these examples, the derivation ratio is generally called a *probability*: probability of death, of divorce; lethal probability.

Here also the probability of the event increases with the interval elapsed, and so the probability of death or divorce in two years after the birth or marriage is obviously greater than that in the first year, but if the implied event is a necessary but not a sufficient condition of the derived event, we may also speak of the probability of the derived event without limitation of time. Thus we speak of the divorce probability of a marriage or of the lethal probability of an illness, while we cannot speak of the death-probability of a newborn child without limitation of time, because every person who comes into the world is bound to die.

We also speak of the probability of an event if it is susceptible to repetition on the same person, for instance, the marriage probability of a person, and this also may take into account not only first marriages but also subsequent ones.

But when the cases of the derived events are more numerous than the implied ones, as in the case of legitimate births to marriages, we no longer speak of a probability. The ratio is in this case called a *repetition ratio*: it gives rise to a concept different from that of the derived, as well as that of the implied phenomenon, and we shall speak of it later among the ratios that can be solved.

We can also think of derived ratios in which both the derived and the implied phenomena are static in character, for example, the ratios of divorced persons to all persons who are or have been married, or the ratio of sick people to population. These ratios must be considered as composition ratios: usually, however, the two phenomena which enter into the composition ratio start at the same time (e.g. male births per hundred births), while in a derivation ratio the implied phenomenon starts before the derived one.

Lastly, we may consider the derivation ratio in which the derived phenomenon is static in character and the implied one is an event. In this case, the latter is the event from which the static condition started: such is, for example, the ratio of divorced people to the number of divorces, and of inhabitants to

births, where it is necessary to state the interval to which the event refers (e.g. one year or one month). These ratios give rise to a concept of duration which is different to that of both the derived and the implied phenomena, and we shall consider them in later paragraphs when speaking of ratios which can be solved.

V.4. RATIOS THAT CAN BE SOLVED

V.4.I. DURATION RATIOS

Ratios that can be solved give rise, as has been said, to a new concept different from each one of the terms. Such ratios are divided into two categories: *duration ratios* and *repetition ratios*, to which reference has already been made.

Duration ratios relate the consistency of a phenomenon to the amount of its periodical renewal. The quotient that is obtained shows the duration of the phenomenon. If, for example, we divide the number of employees of a firm (supposing the number constant) by the number of new employees engaged in a certain interval (again supposing this number constant), we obtain the average length of stay of each employee in the firm.

It is easy to see the value of such a ratio. Suppose a firm has 150 employees, while the new employees or the employees leaving the firm (owing to death, dismissal, end of contract, etc.) number 10 per year. On such a hypothesis the personnel of the firm is renewed every year by $1/15$ so that it will take 15 years for it to be completely renewed; this is equivalent to saying that every employee stays with the firm for an average of 15 years. Thus, in order to obtain the average duration of stay of an employee in the firm, it is necessary to calculate the ratio $150 : 10 = 15$.

But in reality the phenomena that have to be compared in order to obtain the duration are not constant but vary with time. In such a case, in order to obtain, for example, the average duration of stay of an employee in a firm, we can divide the number of employees in the firm in any year by the number of new employees, not indeed of the same year, but of a year previous by a period roughly equal to what may be considered to be the half of the average duration of stay.

Obviously it is supposed that the average duration of stay is not known; otherwise there would be no need to calculate the duration ratio, but we may have a rough idea of it, and the hypothetical value can be used in order to calculate a first approximation of the duration ratio; this first approximation may then be used in order to obtain a second approximation, and so on until the new approximation obtained does not materially differ from the preceding one.

It should also be noticed that, instead of calculating the ratio between the total number of employees and the number of new employees, or more generally between the consistency of the phenomenon and its inward movement, one might equally legitimately calculate the ratio between the consistency of the phenomenon and the outward movement; in our case between the total number of employess and the number of employees leaving the firm. When the inward and outward movements differ materially, one can calculate the ratio between the consistency of the phenomenon and half the sum of the inward and outward movements, or the mean of the two ratios obtained by dividing the consistency of the phenomenon by the inward movement and the outward movement.

There are phenomena which are renewed by more than one inward and outward movement. In such cases, when determining the average duration, one must take into account all the movements which play a part in renewing the phenomenon.

V.4.2. REPETITION RATIOS

Another category of ratios that can be solved are *repetition ratios*, which show, for example, how many times on the average a person who bequeaths an inheritance has himself inherited during his lifetime. If the average number of inheritances in a country is 200,000 per year and the number of hereditary shares is 500,000 and provided that these numbers remain roughly constant through time, one can say that, on the average, any given person has during his lifetime inherited a number of times equal to $500,000/200,000 = 2.5$.

In fact the difference between the number of inheritances bequeathed and the number of hereditary shares received, cannot

be explained by the fact a number of those who receive shares consume them during their lifetime and thus do not bequeath inheritances. Even though this may happen in certain cases, such cases are balanced by the fact that there will also be persons who have not inherited, but who bequeath inheritances amassed through their own work. The difference is, at least for the most part, due to the fact that the majority of the people inheriting figure as such more than once.

In the same way, if we divide the number of legitimate births by the number of marriages, we obtain the average number of births per family.

A great advantage of repetition and duration ratios is that they supply results that the basic statistics do not directly give. For example, the latter do not always tell us the number of children per family nor the average number of times that each person inherits, but we can obtain this information by means of said repetition ratios.

In this case too, the results would however be exact only if both the terms of the ratio (inheritances and shares, legitimate births and marriages) remain approximately constant through time. This is not usually the case. It will then be advisable to divide the number of the derived event by the number of the implied event referred to a past date, preceding the date to which the derived event refers by an interval corresponding to an appropriate mean of the intervals between the occurring of the implied event (marriage, inheritance) and that of the derived event (births, shares). For marriages and births, for example, we may divide the births of one year by the marriages of the preceding year, if as a mean we choose the mode; or for the average interval between the marriage and the births of successive children (which may be six or seven or eight years according to the country) if we choose the arithmetic means.

These methods, however, do not give reliable results when large variations occur in the annual number of births and marriages, as, for example, is often the case during wars or in the post-war years.

Sometimes, however, the statistics allow us to apply more accurate methods in determining the repetition ratios. For example, with reference to the calculation of the average number

of children per marriage, it is worth recalling the method proposed and applied on the occasion into an investigation into the number of legitimate live births in Italy in 1927 (1).

This investigation enables us to determine how many of the legitimate live births in 1927 came from marriages celebrated in that year and how many from those celebrated in each previous years as far back as 1891, the most distant year contributing to legitimate live births in 1927.

In order to obtain such information, a necessary preliminary stage is the elimination, from the number of births in 1927, of the influence of the different number of marriages ascertained from 1927 to 1891. For this purpose it is only necessary to multiply the number of legitimate live births in the year 1927 coming from marriages celebrated in the year x by the ratio between the number of marriages celebrated in 1927 and that of the marriages celebrated in the year x .

In this way we can determine the relative frequency with which the live births in 1927 were derived from marriages in the same year or from those in each of the previous years on the hypothesis that the number of marriages was the same for all the years.

These relative frequencies, freed from the influence of disturbing factors (limited number of cases, influence of war and post-war conditions on the fertility of marriages) by applying suitable methods which we shall discuss in subsequent chapters, are then multiplied by the corresponding number of marriages celebrated in the year under consideration and in each of the previous years up to the earliest year from which the births of the year under consideration can come.

The sum of these products constitutes the weighted average of the number of marriages during the period 1891-1927, in which the percentages of legitimate live births in 1927 coming from marriages in the previous single years have been used as

(1) C. GINI - *Di un procedimento per la determinazione del numero medio dei figli legittimi per matrimonio* (On a method for determining the average number of legitimate children per marriage), «Metron» vol. X, nos. 1/2, 1952; French abstract in «Revue de l'Institut de Statistique» Année I. Liv. I. 1933.

weights. It gives the *virtual number of marriages for the legitimate live births* occurring in 1927, and represents the number of marriages to which the number of legitimate live births in 1927 must be related in order that the average number of legitimate live births per marriage may be obtained.

In the following table, taken from the above-cited article of Gini, are grouped the results obtained by applying the process set out above, for the years 1903 to 1930.

Year	Legitimate live births per marriage	Year	Legitimate live births per marriage
1930 (1)	3.51	1916	3.54
1929	3.32	1915	4.22
1928	3.42	1914	4.16
1927	3.50	1913	4.21
1926	3.51	1912	4.28
1925	3.55	1911	4.14
1924	3.80	1910	4.36
1923	3.73	1909	4.29
1922	3.84	1908	4.45
1921	3.95	1907	4.19
1920	4.65	1906	4.27
1919	3.54	1905	4.37
1918	2.99	1904	4.42
1917	3.70	1903	4.26

(1) The data from 1924 to 1930 refer to post-war territory, those before 1924 to pre-war territory.

LYDIA SPAVENTA DE NOVELLIS

La legge di Tait e l'evoluzione della fecondità legittima attraverso il tempo

La legge di Tait, secondo la quale « la fecondità delle donne coniugate è proporzionale alla differenza tra 50 anni e la loro età », fu ispirata al suo autore dall'osservazione delle tabelle nelle quali il dott. Duncan aveva raccolto i dati relativi alle 16.385 nascite legittime secondo l'età della madre, avvenute nel 1855 nelle città di Edinburgo e Glasgow (1).

Il Tait, osservando il numero medio dei nati in un anno da un gruppo di donne coniugate di una data età, lo trovò all'incirca proporzionale alla differenza tra 50 anni e quell'età, ed espresse la legge con l'equazione lineare

$$f_x = K (50 - x) \quad (1)$$

in cui f_x è la percentuale di donne coniugate di età x che divengono madri entro l'anno e K un coefficiente di fecondità da determinare di volta in volta, e che sui dati osservati egli calcolò pari a 1,5. Più tardi lo Yule, applicando il metodo di Tait, determinò i coefficienti di fecondità per ciascuna classe di donne

(1) V. J.M. DUNCAN, *Fecundity, Fertility, Sterility and allied Topics*, Edinburg, Black, 1866, pag. 241 e segg. e in particolare la Part VI: *Note on Formulae representing the Fecundity and Fertility of Women by prof. Tait*, già pubblicata in «Transactions of the Royal Society of Edinburg». Va notato che il Duncan, per studiare la fecondità alle varie età, aveva limitato la sua osservazione alle nascite legittime, che in questo caso rappresentavano il 90% di tutte le nascite, perchè riteneva che i dati riferentisi a queste fossero più attendibili e completi.

maritate in Inghilterra e in base ad essi trovò i coefficienti di fecondità totale potenziale (1).

Il coefficiente di fecondità del Tait, che per l'età di 50 anni risulta nullo, è il parametro angolare di una retta discendente, la cui inclinazione esprime l'andamento della fecondità legittima. Per poter calcolare il valore di K e quindi tracciare la retta con la formula di Tait, occorre avere i dati sui nati distribuiti secondo l'età della madre.

Il Gini ne ha corretta la formula con un'altra, che consente di calcolare i valori di K anche senza conoscere la distribuzione dei nati secondo l'età della madre :

$$K = \frac{100 Nl}{Dc (50-x)}$$

Il K cioè risulta dal rapporto tra il numero complessivo dei nati legittimi (Nl) e il numero delle donne coniugate (Dc) moltiplicato per la differenza tra 50 anni e la loro età media (x).

Considerando l'età feconda tra i 15 e i 49 anni, si raggruppano le donne in classi quinquennali di età e se ne calcola l'età media facendo la semisomma delle età limite di ciascun gruppo. I gruppi comprendono le donne dai 15 ai 19 anni compiuti, dai 20 ai 24, dai 25 ai 29, dai 30 ai 34, dai 35 ai 39, dai 40 ai 44, dai 45 ai 49. L'età media sarà rispettivamente di 17,50, 22,50, 27,50, 32,50, 37,50, 42,50, 47,50.

E qui il Gini ha apportato un'altra correzione, allo scopo di rendere i dati più aderenti alla realtà del fenomeno al quale si riferiscono.

Essendo infatti la fecondità per 15, 16, 17 anni assai bassa, tendente a crescere verso i 20, tra 15 e 20 anni si è considerata età media quella di 19. E anche per il secondo gruppo, essendo la fecondità maggiore spostata più verso i 25 anni che verso i 20, si è assunta come età media quella di 23 anni (2).

(1) V. U. YULE, *On the changes in the marriage and birth-rates in England and Wales during the past half century, with an inquiry as to their probable causes*, in « Journal de la Société de Statistique de Paris, » 1906.

(2) Vedi in proposito C. GINI: *La dinamica della popolazione*, in « Demografia », vol. XVIII del trattato d'Igiene, pubblicato sotto la direzione del Prof. O. Casagrande, Torino 1930, pag. 710 e segg.; C. GINI: *Sur la mesure de la fécondité des mariages*, comunicazione presentata alla XXI Sessione dell'Istituto Internazionale di Statistica, Messico 1933 e pubblicata in « Bull. de l'Inst. Int. de Statistique », Tome XXVII, 2^{ème} livraison, La Haye, 1934, pag. 76.

La legge enunciata dal Tait è stata spesso giudicata del tutto arbitraria e l'indice che se ne ricava grossolano e non significativo. Tuttavia non si è mai tentato di giungere a una visione ampia dell'andamento dell'indice di Tait nel tempo e nello spazio, ma si sono fatti calcoli per brevi periodi e per spazi limitati (1).

È questo lo scopo del presente lavoro, nel quale il calcolo dell'indice di Tait è stato fatto per molti paesi e per lunghi periodi di tempo accanto a quello del quoziente di fecondità legittima e del quoziente di natalità. Il confronto tra l'andamento degli indici può offrire materia ad interessanti osservazioni. La formula adottata è stata quella del Gini, come pure la correzione per l'età media.

Gran parte del materiale su cui è stato calcolato l'indice era già raccolto presso l'Istituto di Statistica ed è stato aggiornato con i dati degli ultimi censimenti e delle più recenti rilevazioni statistiche a disposizione. Naturalmente la lunga parentesi bellica ha causato quasi sempre interruzioni più o meno ampie in tali rilevazioni, e in alcuni casi per situazioni politiche e per modificazioni territoriali non definite, non si è ancora tornati alla normalità.

Inoltre si tende ormai alla abolizione della distinzione tra fecondità legittima e illegittima, e gli stessi annuari demografici dell'ONU non ne tengono conto.

Prima di esporre i risultati occorre fare alcune osservazioni preliminari, che sono di un doppio ordine. Per quanto riguarda la valutazione dei dati di cui si dispone, bisogna dire anzitutto che i criteri di rilevazione non sono identici in tutti i paesi e nemmeno nello stesso paese attraverso il tempo; inoltre non

(1) Il GINI ha calcolato età media, saggio di fecondità e coefficiente K di Tait per le singole regioni del Regno d'Italia all'epoca dei censimenti, fino al censimento del 1921 e si è servito di tale calcolo indiretto per determinare, attraverso i saggi di fecondità delle madri in età feconda, il numero dei nati negli anni futuri.

Un calcolo dei quozienti teorici di fecondità legittima per ciascun anno di età fra 15 e 40 anni con la formula del Tait è stato fatto per il triennio 1935-37 e 1950-52 dallo ZANGARELLI *Le variazioni della fecondità italiana dal 1936 al 1951* in «Statistica» 1959, n. 2.

sempre, e per molti motivi, essi sono ispirati alla stessa rigorosità. Molto spesso le donne separate, anche legalmente, dal coniuge sono comprese tra le coniugate. Per quanto riguarda le unioni consensuali, nei paesi dove tali unioni libere sono numerose, se nei censimenti non vi è indicazione per una classificazione a parte, accade quasi sempre che la denunciante, potendo scegliere tra il dichiararsi nubile o coniugata, preferisca dichiararsi coniugata (1). Quanto poi al numero dei nati non vi è criterio uniforme nella denuncia dei nati vivi, dato che la denuncia spesso non viene fatta lo stesso giorno, e nel computo dei nati vivi sono compresi qualche volta anche quelli che sopravvivono solo poche ore, qualche volta solo quelli che superano il secondo o il terzo giorno dalla nascita. Inoltre l'età della madre è quasi sempre indicata con l'anno di nascita; ma sono non infrequenti i casi di madri la cui età è ignota.

D'altra parte è necessario sottolineare che i dati si riferiscono ad un fenomeno sul cui andamento influiscono molti e diversi fattori.

Anzitutto bisogna ricordare che sulla fecondità di ciascun gruppo esercita grande influenza il numero delle nuove spose, le quali, sia per ragioni di ordine psicologico (desiderio di avere il primo figlio) sia per ragioni di ordine fisiologico, sono più feconde; e che inoltre le nuove spose danno un apporto, che può essere più o meno rilevante, di concepimenti antenuziali, indubbiamente più frequenti per le spose giovani, in particolare per quelle che appartengono al primo raggruppamento (2), vale a dire che sono al disotto dei vent'anni, e in parte anche al secondo.

Inoltre, come è stato più volte messo in evidenza dal Gini, quando si vogliano calcolare i tassi di fecondità matrimo-

(1) La Commissione della Popolazione dell'Organizzazione delle Nazioni Unite raccomanda di classificare le donne unite consensualmente tra le coniugate.

(2) Il PAGLINO nel suo lavoro: *Su la misura della fecondità matrimoniale della donna*, «Metron», vol. IV, n. 2-3-4, 1941, ha trovato che nel periodo 1930-32 sui saggi di fecondità della popolazione generale del Regno di Italia per il gruppo di età 15-19 il rapporto dei parti nei primi nove mesi di matrimonio è di circa il 38% rispetto al totale dei nati.

niale non si può prescindere da altre circostanze che incidono notevolmente sui risultati :

1) Le donne coniugate, a parte i casi di concepimenti prenuziali, sono normalmente esposte al rischio del parto solo nove mesi dopo il matrimonio.

2) Dopo il parto, tenendo conto di tutte le circostanze di ordine fisiologico, le donne rimangono escluse in media per un anno dal rischio di un parto successivo.

3) Le nuove vedove, le divorziate, e le separate legalmente restano esposte ancora per nove mesi al rischio del parto.

Va ancora detto che l'andamento della fecondità e della natalità presenta qua e là nella linea discendente delle pause e delle riprese, che sono connesse spesso con fattori esterni di perturbazione, primo fra tutti la guerra che divide per più o meno lunghi periodi di tempo i coniugi, mutila del capo alcune famiglie, mentre in quelle nelle quali il coniuge rientra, provoca nell'anno successivo alla sua fine, un aumento delle nascite per la ripresa della normale vita coniugale. Questo si ripercuote anche a distanza di anni, quando giungono al massimo di fecondità le generazioni nate nei periodi di aumentata natalità per cause varie.

Per tener conto di tutti questi fattori di perturbazione, bisognerebbe dunque calcolare molteplici saggi di eliminazione, e per far ciò si dovrebbe disporre di una quantità di elementi che a volte mancano. Quand'anche poi si avessero tutti questi elementi, i calcoli sarebbero necessariamente lunghissimi e quindi realizzabili solo per periodi limitati di tempo e per spazi ristretti.

Ne sono esempi i lavori condotti dal Mukherji per la Francia per il solo periodo 1926-30 (1), dal Paglino per la popolazione del Regno d'Italia per il periodo 1931-35 (2), dal Luivni per la Svizzera per il quinquennio 1931-1935 (3). In questi lavori gli autori, approfondendo gli argomenti, sono potuti giun-

(1) MUKHERJI A., *Etude statistique de la fécondité matrimoniale*, Paris, 1935.

(2) PAGLINO F., *Su la misura della fecondità matrimoniale della donna*, in «Metron», vol. XIV, 1940-41.

(3) LUVINI A., *La fecondità matrimoniale in Svizzera*, in «Genus», vol. IX 1950-52.

gere a risultati che indubbiamente riproducono i fenomeni studiati con grande aderenza alla realtà, e che offrono, per la molteplicità dei calcoli eseguiti, ampio materiale di comparazione e di osservazione.

* * *

Nel presente lavoro sono stati calcolati natalità, fecondità legittima, età media e indice di Tait per le coniugate tra i 15 e i 49 anni per i seguenti paesi:

Finlandia, dal 1880 al 1950 ogni dieci anni.

Svezia, dal 1870 al 1930 ogni dieci anni, e poi per gli anni 1935, 1940, 1945, 1950.

Norvegia, per gli anni 1855, 1865, 1875, 1891, 1900, 1910, 1920, 1930, 1946, 1950.

Danimarca, dal 1850 al 1930 ogni dieci anni e poi per gli anni 1935, 1940, 1945, 1948, 1950.

Islanda, per gli anni 1930, 1940 e 1950.

Estonia, per gli anni 1922 e 1934.

Lettonia, per gli anni 1920, 1925, 1930, 1935.

Inghilterra e Galles, dal 1851 ogni dieci anni fino al 1931, e poi per gli anni 1940, 1945, 1950.

Irlanda (Eire), dal 1871 al 1911 ogni dieci anni e poi per gli anni 1926, 1936, 1946.

Scozia, per gli anni dal 1861 al 1951 ogni dieci anni.

Belgio, per gli anni 1846, 1856, 1866 e poi dal 1880 ogni dieci anni fino al 1930 e infine per il 1947.

Olanda, dal 1849 al 1909 ogni dieci anni e poi per gli anni 1920, 1930, 1947.

Lussemburgo, per gli anni 1922, 1927, 1930, 1935, 1947.

Germania, per gli anni 1871, 1880, 1890, 1900, 1910, 1925, 1933, 1946, 1950.

Austria, dal 1880 al 1920 ogni dieci anni e poi per il 1934 e per il 1951.

Serbia, per gli anni 1895, 1900.

Cecoslovacchia, per gli anni 1921 e 1930.

Ungheria, per gli anni 1900, 1910, 1920, 1930, 1941.

Bulgaria, per gli anni 1900, 1910, 1920, 1926, 1934.

Svizzera, dal 1860 al 1930 ogni dieci anni (solo per il 1888 sono fatti i calcoli invece che per il 1890) e poi per il 1941.

Francia, dall'anno 1851 al 1921 ogni dieci anni e poi per gli anni 1926, 1931, 1936, 1946.

Spagna, per gli anni 1887, 1900, 1910, 1920, 1940.

Portogallo, per gli anni 1864, 1890, 1900, 1911, 1920, 1930, 1940, 1950.

Italia, per gli anni dal 1881 al 1931 ogni dieci anni e per il 1936 e per il 1951.

Grecia, per gli anni 1870, 1879, 1907, 1920, 1928.

Sono stati inoltre esaminati i dati per i seguenti paesi extra-europei :

Canada, per gli anni 1921, 1931, 1941, 1951.

Cile, per gli anni 1907, 1920, 1930.

Giappone, per gli anni 1898, 1908, 1918, 1925, 1950.

Nuova Zelanda, per gli anni 1881, 1901, 1911, 1921, 1926, 1936, 1945, 1951.

Confederazione Australiana, per gli anni dal 1891 al 1921 ogni dieci anni e per gli anni 1933, 1947.

Unione Sud Africa, Libero Stato d'Orange, Transvaal, Capo di Buona Speranza, Natal, per gli anni 1911, 1918, 1921, 1926, 1936.

Per i paesi i cui dati a disposizione lo hanno consentito, sono stati calcolati i numeri indici, prendendo generalmente come anni base il 1900 o gli anni vicini, dato che l'andamento della natalità e della fecondità appaiono lievemente decrescenti fino alla fine del secolo e solo da quest'epoca cominciano a diminuire decisamente.

ANNI	VALORI EFFETTIVI				NUMERI INDICI				MEDIA DEI NATI NEGLI ANNI
	Natalità	Saggio fecond. legitt.	Indice K	Età media delle coniug.	Natalità	Saggio fecond. legitt.	Indice K	Età media delle coniug.	

FINLANDIA (1900 = 1)

1880 . .	35.92	259.26	1.733	35.043	1,12	1,05	1,04	0,99	1880—81—82
1890 . .	32.91	246.57	1.701	35.506	1,02	1,00	1,02	1,01	1890—91—92
1900 . .	32.12	247.00	1.671	35.218	1,00	1,00	1,00	1,00	1900—01—02
1910 . .	29.72	234.52	1.585	35.206	0,93	0,95	0,95	1,00	1910—11—12
1920 . .	24.33	202.90	1.440	35.944	0,76	0,82	0,86	1,02	1920—21—22
1930 . .	22.28	173.52	1.188	35.402	0,69	0,70	0,71	1,01	1930—31—32
1940 . .	21.10	148.60	1.018	35.406	0,66	0,60	0,61	1,01	1939—40—41
1950 . .	24.50	154.82	1.057	35.358	0,76	0,62	0,63	1,00	Dati ricavati dall'annuario demografico delle N.U.

SVEZIA (1900 = 1)

1870 . .	29.75	229.41	1.784	37.144	1,11	1,03	1,07	1,02	1870—71—72
1880 . .	29.26	236.35	1.769	36.638	1,09	1,06	1,07	1,00	1880—81—82
1890 . .	27.73	228.58	1.701	36.562	1,03	1,03	1,02	1,00	1890—91—92
1900 . .	26.84	222.69	1.660	36.588	1,00	1,00	1,00	1,00	1900—01—02
1910 . .	24.01	193.88	1.413	36.283	0,89	0,87	0,85	0,99	1910—11—12
1920 . .	22.92	165.01	1.191	36.142	0,85	0,74	0,72	0,99	1920—21—22
1930 . .	14.93	103.59	0.758	36.332	0,56	0,47	0,46	0,99	1930—31—32
1935 . .	13.80	93.44	0.679	36.256	0,51	0,42	0,41	0,99	1934—35—36
1940 . .	15.10	95.11	0.675	35.907	0,56	0,43	0,41	0,98	1939—40—41
1945 . .	20.40	118.20	0.833	35.804	0,76	0,53	0,50	0,98	1944—45—46
1950 . .	16.40	94.33	0.675	36.105	0,61	0,42	0,41	0,99	1950

NORVEGIA (1900 = 1)

1855 . .	32.80	268.02	1.843	35.456	1,11	1,07	1,04	0,99	1855—56—57
1865 . .	31.25	247.34	1.800	36.257	1,06	0,99	1,02	1,01	1865—66—67
1875 . .	31.77	259.59	1.681	34.558	1,08	1,04	0,95	0,96	1875—76—77
1891 . .	30.27	257.97	1.856	36.102	1,02	1,03	1,05	1,01	1890—91—92
1900 . .	29.54	249.57	1.770	35.902	1,00	1,00	1,00	1,00	1900—01—02
1910 . .	25.82	224.01	1.591	35.925	0,87	0,90	0,90	1,00	1910—11—12
1920 . .	24.87	206.76	1.430	35.547	0,84	0,83	0,81	0,99	1920—21—2
1930 . .	17.07	136.15	0.973	36.008	0,85	0,54	0,55	1,00	1929—30—2
1946 . .	22.60	147.20	1.059	36.102	0,76	0,58	0,59	1,00	
1950 . .	19.10	121.44	0.872	36.060	0,64	0,48	0,49	1,00	Dati ricavati dall'annuario demografico delle N.U.

ANNI	VALORI EFFETTIVI				NUMERI INDICI				MEDIA DEI NATI NEGLI ANNI
	Natalità	Saggio fecond. legitt.	Indice K	Età media delle coniug.	Natalità	Saggio fecond. legitt.	Indice K	Età media delle coniug.	

DANIMARCA (1900 = 1)

1850 . .	31.80	233.25	1.713	36.386	1,07	1,06	1,23	1,06	1850—51—51
1860 . .	32.80	231.76	1.700	36.366	1,11	1,06	1,22	1,06	1859—60—61
1870 . .	30.06	218.47	1.653	36.781	1,01	1,00	1,07	1,07	1869—70—71
1880 . .	32.03	241.03	1.741	36.158	1,08	1,10	1,25	1,05	1779—80—81
1890 . .	30.95	233.91	1.677	36.052	1,04	1,07	1,30	1,05	1889—90—91
1900 . .	29.69	219.09	1.395	34.300	1,00	1,00	1,00	1,00	1900—01—02
1910 . .	27.07	192.93	1.323	35.422	0,91	0,88	0,95	1,03	1910—11—12
1920 . .	23.56	161.16	1.114	35.536	0,79	0,74	0,80	1,04	1920—21—22
1930 . .	18.39	117.57	0.822	35.609	0,62	0,54	0,59	1,04	1929—30—31
1935 . .	17.70	109.16	0.753	35.698	0,60	0,50	0,54	1,04	1934—35—36
1940 . .	18.30	106.87	0.731	35.389	0,62	0,49	0,52	1,03	1939—40—41
1945 . .	23.50	132.04	0.915	35.573	0,79	0,60	0,66	1,04	1944—45—46
1948 . .	20.30	116.49	0.799	35.525	0,68	0,53	0,57	1,04	1947—48—49
1950 . .	18.60	104.82	0.728	35.608	0,63	0,48	0,52	1,04	1949—50—51

ISLANDA (1 = 0861)

1930 . .	24.3	207.32	1.419	35.393	1,00	1,00	1,00	1,00	1930—31—32
1940 . .	20.9	144.26	1.087	36.731	0,86	0,69	0,76	1,03	1939—40—41
1950 . .	28.7	169.21	1.115	34.821	1,18	0,81	0,78	0,98	Dati ricavati dall'an- nuario demografico delle N.U.

ESTONIA (1922 = 1)

1922 . .	20.04	123.65	0.837	35.839	1,00	1,00	1,00	1,00	1922
1934 . .	15.04	100.70	0.713	35.879	0,75	0,81	0,85	1,00	1933—34—35

LETTONIA (1920 = 1)

1920 . .	22.35	173.76	1.306	36.700	1,00	1,00	1,00	1,00	1920—21—22
1925 . .	22.33	158.75	1.122	35.850	1,00	0,91	0,85	0,97	1924—25—26
1930 . .	19.38	132.52	0.930	35.748	0,86	0,76	0,71	0,97	1929—30—31
1935 . .	17.65	115.20	0.711	33.913	0,78	0,66	0,54	0,92	1934—35—36

INGHILTERRA E GALLES (1901 = 1)

1851 . .	34.09	247.14	1.587	34.426	1,19	1,21	1,16	0,98	1850—51—52
1861 . .	34.77	244.27	1.572	34.465	1,21	1,20	1,15	0,98	1860—61—62
1871 . .	35.46	253.12	1.660	34.754	1,24	1,24	1,22	0,99	1870—71—72
1881 . .	34.06	248.47	1.634	34.800	1,19	1,22	1,20	0,99	1880—81—82
1891 . .	30.83	228.02	1.520	35.000	1,08	1,12	1,11	1,00	1890—91—92
1901 . .	28.67	203.81	1.364	35.062	1,00	1,00	1,00	1,00	1900—01—02
1911 . .	24.50	168.75	1.181	35.709	0,85	0,83	0,87	1,02	1910—11—12
1921 . .	22.76	149.23	1.078	36.161	0,79	0,73	0,79	1,03	1920—21—22
1931 . .	15.81	101.93	0.736	36.157	0,55	0,50	0,54	1,03	1930—31—32
1940 . .	14.90	84.48	0.594	35.774	0,52	0,41	0,44	1,02	1939—40—41
1945 . .	16.70	90.89	0.549	33.442	0,58	0,45	0,40	0,95	1944—45—46
1950 . .	15.80	90.96	0.648	35.968	0,55	0,45	0,48	1,03	1949—50—51

ANNI	VALORI EFFETTIVI				NUMERI INDICI				MEDIA DEI NATI NEGLI ANNI
	Natalità	Saggio fecond. legitt.	Indice K	Età media delle coniug.	Natalità	Saggio fecond. legitt.	Indice K	Età media delle coniug.	

IRLANDA (1901 = 1)

1871 . .	27.74	267.73	1.811	35.216	1,22	1,09	1,04	0,98	1870—71—72
1881 . .	24.26	244.98	1.745	35.960	1,06	1,00	1,00	1,00	1880—81—82
1891 . .	22.50	240.94	1.772	36.401	0,99	0,98	1,01	1,01	1890—91—92
1901 . .	22.79	245.76	1.749	35.947	1,00	1,00	1,00	1,00	1900—01—02
1911 . .	23.14	251.11	1.824	36.236	1,02	1,02	1,04	1,01	1910—11—12
1926 . .	20.56	217.54	1.699	37.195	0,90	0,89	0,97	1,03	1925—26—27
1936 . .	19.41	208.53	1.618	37.116	0,85	0,85	0,93	1,03	1935—36—37
1946 . .	22.30	219.30	1.654	36.738	0,98	0,89	0,95	1,02	1945—46—47

SCOZIA (1901 = 1)

1861 . .	34.80	274.00	1.776	34.576	1,18	1,17	1,12	0,98	1860—61—62
1871 . .	34.75	276.56	1.804	34.667	1,18	1,18	1,14	0,98	1870—71—72
1881 . .	33.63	269.51	1.800	35.027	1,13	1,15	1,14	0,99	1880—81—82
1891 . .	30.85	253.57	1.730	35.340	1,05	1,08	1,09	1,00	1890—91—92
1901 . .	29.51	234.31	1.583	35.202	1,00	1,00	1,00	1,00	1900—01—02
1911 . .	25.81	198.67	1.396	35.774	0,87	0,85	0,88	1,02	1910—11—12
1921 . .	25.59	188.94	1.269	35.115	0,87	0,81	0,80	1,00	1920—21—22
1931 . .	19.12	140.56	1.004	36.111	0,65	0,60	0,63	1,03	1930—31—32
1951 . .	17.70	109.77	0.778	35.908	0,59	0,46	0,49	1,02	Dati ricavati dall'annuario demografico delle N.U.

BELGIO (1900 = 1)

1846 . .	28.88	251.03	1.931	37.002	1,00	1,15	1,32	1,05	1845—46—47
1856 . .	30.78	276.88	2.155	37.152	1,06	1,27	1,47	1,06	1856—57—58
1866 . .	32.09	280.71	2.045	36.275	1,11	1,29	1,40	1,03	1866—67—68
1880 . .	31.24	263.35	1.889	36.061	1,08	1,21	1,29	1,03	1880—81—82
1890 . .	29.13	238.44	1.677	35.778	1,01	1,10	1,15	1,02	1890—91—92
1900 . .	28.92	217.58	1.464	35.137	1,00	1,00	1,00	1,00	1900—01—02
1910 . .	23.16	158.88	1.086	35.376	0,80	0,73	0,74	1,01	1910—11—12
1920 . .	21.90	138.63	0.984	35.906	0,76	0,64	0,67	1,02	1920—21—22
1930 . .	18.32	106.01	0.719	35.252	0,63	0,49	0,49	1,00	1930—31—32
1947 . .	17.80	104.16	0.762	36.327	0,62	0,48	0,52	1,03	1946—47—48

OLANDA (1899 = 1)

1849 . .	34.30	297.02	2.225	36.648	1,07	1,10	1,17	1,03	1849—50—51
1859 . .	34.15	285.88	2.130	36.581	1,07	1,06	1,12	1,02	1859—60—61
1869 . .	35.25	288.01	2.146	36.580	1,10	1,06	1,13	1,02	1869—70—71
1879 . .	35.72	293.00	2.129	36.239	1,12	1,08	1,12	1,01	1879—80—81
1889 . .	33.28	284.50	2.042	36.066	1,04	1,05	1,08	1,01	1889—90—91
1899 . .	31.92	270.44	1.897	35.748	1,00	1,00	1,00	1,00	1899—00—01
1909 . .	28.54	230.39	1.624	35.818	0,89	0,85	1,86	1,00	1909—10—11
1920 . .	27.17	203.00	1.354	35.013	0,85	0,75	0,71	0,98	1920—21—22
1930 . .	22.55	176.05	1.138	35.661	0,71	0,65	0,60	1,00	1929—30—31
1947 . .	27.70	186.78	1.320	35.854	0,87	0,69	0,70	1,00	1946—47—48

ANNI	VALORI EFFETTIVI				NUMERI INDICI				MEDIA DEI NATI NEGLI ANNI
	Natalità	Saggio fecond. legitt.	Indice K	Età media delle coniug.	Natalità	Saggio fecond. legitt.	Indice K	Età media delle coniug.	

LUSSEMBURGO (1922 = 1)

1922 . .	20.36	145.93	1.030	35.831	1,00	1,00	1,00	1,00	1922—23—24
1927 . .	20.57	136.65	0.927	35.278	1,01	0,94	0,90	0,98	1926—27—28
1930 . .	19.62	130.86	0.873	31.011	0,96	0,90	0,85	0,87	1930—31—32
1935 . .	15.21	96.99	0.669	35.633	0,75	0,66	0,65	0,99	1934—35—36
1947 . .	14.80	90.31	0.686	36.880	0,72	0,61	0,66	1,02	Dati ricavati dall'annuario demografico delle N.U.

GERMANIA (1900 = 1)

1871 . .	39.48	286.00	2.018	35.827	1,11	1,11	1,14	1,01	1872
1880 . .	37.29	263.98	1.821	35.505	1,05	1,02	1,02	1,00	1880—81—82
1890 . .	36.15	257.59	1.783	35.555	1,02	1,00	1,00	1,00	1890—91—92
1900 . .	35.47	257.96	1.777	35.486	1,00	1,00	1,00	1,00	1900—01—02
1910 . .	28.88	193.40	1.336	35.325	0,81	0,75	0,75	1,00	1910—11—12
1925 . .	19.98	120.04	0.849	35.867	0,56	0,47	0,48	1,01	1920—21—22
1933 . .	15.94	93.22	0.793	36.277	0,45	0,36	0,45	1,02	1932—33—34
1946 . .	16.70 ¹	88.33	0.687	37.139	0,47	0,34	0,39	1,05	
1950 . .	16.57	94.65	0.725	36.943	0,47	0,37	0,41	1,04	Dati ricavati dall'annuario demografico delle N.U.

AUSTRIA (1900 = 1)

1880 . .	38.08	249.78	1.756	35.774	1,03	1,01	1,06	1,02	1880—81—82
1890 . .	37.00	246.99	1.671	35.222	1,00	1,00	1,01	1,01	1890—91—92
1900 . .	36.96	248.01	1.652	34.989	1,00	1,00	1,00	1,00	1900—01—02
1910 . .	31.97	213.31	1.433	35.117	0,86	0,86	0,87	1,00	1910—11—12
1920 . .	22.99	149.79	1.130	36.740	0,62	0,60	0,68	1,05	1920—21—22
1934 . .	13.64	75.01	0.563	36.676	0,37	0,30	0,34	1,05	1933—34—35
1951 . .	14.80	85.64	0.648	36.793	0,40	0,35	0,39	10,5	

UNGHERIA (1900 = 1)

1900 . .	38.39	217.97	1.341	33.744	1,00	1,00	1,00	1,00	1900—01—02
1910 . .	35.33	200.18	1.231	33.744	0,92	0,92	0,92	1,00	1910—11—12
1920 . .	31.85	186.15	1.153	33.856	0,83	0,85	0,86	1,00	1920—21—22
1930 . .	24.26	136.36	0.857	34.093	0,63	0,63	0,64	1,01	1929—30
1941 . .	13.40	153.94	1.037	35.158	0,34	0,71	0,77	1,04	1940—41

BULGARIA (1900 = 1)

1900 . .	39.51	259.47	1.533	33.077	1,00	1,00	1,00	1,00	1900—01—02
1910 . .	41.20	254.25	1.446	32.417	1,05	0,98	0,94	0,98	1910—11—12
1920 . .	40.52	248.73	1.456	32.920	1,03	0,96	0,95	1,00	1920—21—22
1926 . .	35.43	191.44	1.118	32.869	0,90	0,74	0,73	0,99	1925—26—27
1934 . .	27.79	151.25	0.902	33.235	0,70	0,58	0,59	1,00	1934—35—36

¹⁾ Media della natalità nelle tre zone d'occupazione.

ANNI	VALORI EFFETTIVI				NUMERI INDICI				MEDIA DEI NATI NEGLI ANNI
	Natalità	Saggio fecond. legitt.	Indice K	Età media delle coniug.	Natalità	Saggio fecond. legitt.	Indice K	Età media delle coniug.	

SVIZZERA (1900 = 1)

1860 . .	30.76	253.27	1.778	35.759	1,07	1,10	1,11	1,01	1860—61—62
1870 . .	29.51	240.09	1.736	35.713	1,02	1,04	1,09	1,00	1870—71—72
1880 . .	29.46	236.17	1.696	36.075	1,02	1,02	1,06	1,01	1880—81—82
1888 . .	27.16	224.66	1.623	36.155	0,94	0,97	10,2	1,02	1888—89—90
1900 . .	28.81	230.67	1.599	35.574	1,00	1,00	1,00	1,00	1900—01—02
1910 . .	24.45	186.68	1.318	35.831	0,85	0,81	0,82	1,01	1910—11—12
1920 . .	20.45	144.21	1.088	36.744	0,71	0,63	0,68	1,03	1920—21—22
1930 . .	16.98	122.33	0.892	36.290	0,59	0,53	0,56	1,02	1931—32
1941 . .	15.40	157.49	1.025	34.640	0,53	0,68	0,64	0,97	1940—41—42

FRANCIA (1901 = 1)

1851 . .	26.93	182.43	1.280	35.748	1,24	1,35	1,43	1,02	1850—51—52
1861 . .	26.37	172.48	1.194	35.564	1,22	1,28	1,33	1,02	1860—61—62
1871 . .	25.28	164.83	1.111	35.172	1,17	1,22	1,24	1,01	1871—72—73
1881 . .	24.71	165.35	1.119	35.221	1,14	1,23	1,25	0,99	1880—81—82
1891 . .	22.26	146.38	0.992	35.241	1,03	1,09	1,10	1,01	1890—91—92
1901 . .	21.64	134.85	0.898	34.985	1,00	1,00	1,00	1,00	1900—01—02
1911 . .	19.27	113.45	0.759	35.051	0,89	0,84	0,85	1,00	1910—11—12
1921 . .	20.44	136.61	0.948	35.592	0,94	1,01	1,06	1,02	1920—21—22
1926 . .	18.61	117.82	0.792	35.130	0,86	0,87	0,88	1,00	1925—26—27
1931 . .	17.60	101.46	0.673	34.935	0,81	0,75	0,75	1,00	1930—31—32
1936 . .	15.30	87.60	0.566	35.056	0,71	0,64	0,63	1,00	1935—36—37
1946 ¹	20.90	124.45	0.910	36.332	0,91	0,92	1,01	1,03	

SPAGNA (1900 = 1)

1887 . .	36.38	233.71	1.547	34.894	1,05	1,01	1,00	0,99	1886—87—88
1900 . .	34.77	230.91	1.549	35.092	1,00	1,00	1,00	1,00	1900—01—02
1910 . .	31.86	218.09	1.480	35.266	0,92	0,94	0,96	1,00	1910—11—12
1920 . .	30.31	211.00	1.457	35.516	0,87	0,91	0,94	1,01	1920—21—22
1940 . .	22.00	154.66	1.104	35.986	0,63	0,67	0,71	1,02	1935—40—41

PORTOGALLO (1900 = 1)

1864 . .	32.20	230.75	1.647	34.990	1,05	1,00	1,07	1,03	1862
1890 . .	31.93	236.24	1.612	35.348	1,04	1,03	1,05	1,01	1890—91—92
1900 . .	30.56	229.88	1.542	35.090	1,00	1,00	1,00	1,00	1900—01—02
1911 . .	35.09	225.36	1.494	34.920	1,15	0,98	0,97	1,00	1911
1920 . .	32.05	233.74	1.584	35.245	1,05	1,02	1,03	1,00	1920—21—22
1930 . .	30.02	199.97	1.318	34.988	0,98	0,87	0,85	1,00	1930—31—32
1940 . .	27.10	163.55	1.107	35.224	0,89	0,71	0,72	1,00	1939—40—41
1950 . .	24.40	156.64	1.064	35.285	0,80	0,68	0,69	1,01	1949—50—51

¹) Senza la Sarre.

ANNI	VALORI EFFETTIVI				NUMERI INDICI				MEDIA DEI NATI NEGLI ANNI
	Natalità	Saggio fecond. legitt.	Indice K	Età media delle coniug.	Natalità	Saggio fecond. legitt.	Indice K	Età media delle coniug.	

ITALIA (1901 = 1)

1881 . .	37.63	250.31	1.584	34.200	1,14	1,08	1,06	1,00	1880—81—82
1901 . .	32.93	231.74	1.481	34.352	1,00	1,00	1,00	1,00	1900—01—02
1911 . .	31.36	228.20	1.496	34.744	0,95	0,98	1,01	1,01	1910—11—12
1921 . .	30.23	220.29	1.489	35.204	0,92	0,95	1,01	1,02	1920—21—22
1931 . .	25.18	171.71	1.056	33.738	0,76	0,74	0,71	0,98	1930—31—32
1936 . .	23.18	157.83	1.069	35.229	0,70	0,68	0,72	1,03	1935—36—37
1951 ¹	18.34	153.09	1.145	36.631	0,55	0,66	0,77	1,06	1950—51—52

GRECIA (1920 = 1)

1870 . .	28.43	190.85	1.060	32.000	1,49	1,18	0,97	0,91	1869—70—71
1879 . .	25.34	173.58	1.027	33.105	1,33	1,20	0,94	0,94	1878—79—80
1920 . .	19.08	161.46	1.089	35.172	1,00	1,00	1,00	1,00	1921—22
1928 . .	29.42	199.18	1.118	32.196	1,54	1,23	1,03	0,92	1927—28—29

CANADÀ (1921 = 1)

1921 . .	26.00	129.72	0.840	34.550	1,00	1,00	1,00	1,00	1920—21—22
1931 . .	23.32	161.83	1.093	35.194	0,90	1,25	1,30	1,02	1930—31—32
1941 . .	22.20	150.13	0.987	34.787	0,85	1,16	1,18	1,01	
1951 . .	27.20	152.67	0.982	34.447	1,04	1,18	1,17	1,00	

CILE (1907 = 1)

1907 . .	38.56	224.84	1.361	33.479	1,00	1,00	1,00	1,00	1906—07—08
1920 . .	39.02	230.29	1.392	33.454	1,01	1,02	1,02	1,00	1920—21—22
1930 . .	38.48	222.97	1.341	33.384	1,00	0,99	0,99	1,00	1929—30—31

GIAPPONE (1908 = 1)

1898 . .	31.45	202.00	1.241	33.729	0,93	0,87	0,86	1,00	1898—99—1900
1908 . .	33.87	232.31	1.438	33.846	1,00	1,00	1,00	1,00	1908—09—10
1918 . .	32.13	228.88	1.396	33.608	0,95	0,99	0,97	0,99	1918—19—20
1925 . .	33.50	199.30	1.198	33.358	0,99	0,86	0,83	0,99	1924—25—26
1950 . .	28.20	182.84	1.162	34.270	0,83	0,78	0,80	1,01	Dati ricavati dall'annuario demografico delle N.U.

NUOVA ZELANDA (1901 = 1)

1881 . .	38.67	295.63	1.823	33.782	1,49	1,39	1,31	0,97	1880—81—82
1901 . .	25.94	212.62	1.389	34.693	1,00	1,00	1,00	1,00	1900—01—02
1911 . .	26.21	188.11	1.234	34.754	1,01	0,88	0,89	1,00	1910—11—12
1921 . .	24.35	159.14	1.121	35.803	0,94	0,75	0,81	1,03	1920—21—22
1926 . .	21.12	137.05	0.986	36.097	0,81	0,64	0,71	1,04	1925—26—27
1936 . .	16.72	113.01	0.809	36.039	0,64	0,53	0,58	1,04	
1945 . .	23.20	148.03	1.022	35.513	0,89	0,70	0,64	1,02	
1951 . .	24.40	138.71	0.931	35.109	0,94	0,65	0,67	1,01	

1) Non compreso il territorio di Trieste.

ANNI	VALORI EFFETTIVI				NUMERI INDICI				MEDIA DEI NATI NEGLI ANNI
	Natalità	Saggio fecond. legitt.	Indice K	Età media delle coniug.	Natalità	Saggio fecond. legitt.	Indice K	Età media delle coniug.	

CONFEDERAZIONE AUSTRALIANA (1901 = 1)

1891 . .	34.14	270.56	1.632	33.419	1,26	1,29	1,19	0,96	1890—91—92
1901 . .	26.98	209.70	1.370	34.689	1,00	1,00	1,00	1,00	1900—01—02
1911 . .	27.84	204.22	1.359	34.974	1,03	0,97	0,99	1,00	1910—11—12
1921 . .	25.15	171.57	1.141	34.966	0,93	0,82	0,83	1,00	1920—21—22
1933 . .	16.68	111.19	0.786	35.863	0,62	0,53	0,57	1,03	1932—33—34
1947 . .	24.10	137.16	0.909	34.917	0,89	0,65	0,66	1,01	1946—47—48

UNIONE SUD AFRICA (1911 = 1)

1911 . .	32.601	224.880	1.372	33.617	1,00	1,00	1,00	1,00	1911
1918 . .	28.609	187.808	1.191	34.234	0,88	0,84	0,87	0,99	1920—21—22
1921 . .	28.426	191.190	1.225	34.400	0,87	0,81	0,89	1,02	1925—26—27
1926 . .	26.169	177.290	1.162	34.746	0,80	0,79	0,85	1,03	
1936 . .	24.490	156.952	1.003	34.361	0,75	0,70	1,02	1,02	

LIBERO STATO D'ORANGE (1911 = 1)

1911 . .	29.171	201.118	1.182	32.998	1,00	1,00	1,00	1,00	1911
1918 . .	26.772	174.372	1.070	33.712	0,92	0,87	0,91	0,99	1920—21—22
1921 . .	26.877	179.373	1.123	34.034	0,92	0,89	0,95	1,03	1925—26—27
1926 . .	25.982	173.192	1.121	34.552	0,89	0,86	0,95	1,05	
1936 . .	24.030	150.961	0.986	34.699	0,82	0,75	0,83	1,05	

TRANSVAAL (1911 = 1)

1911 . .	37.158	239.522	1.409	33.009	1,00	1,00	1,00	1,00	1911
1918 . .	31.240	192.037	1.188	33.847	0,84	0,80	0,84	1,03	1910—11—12
1921 . .	30.494	194.859	1.224	34.092	0,82	0,81	0,87	1,03	1925—26—27
1926 . .	27.221	178.606	1.151	34.496	0,73	0,75	0,82	1,05	
1936 . .	27.257	166.685	1.017	33.626	0,73	0,70	0,78	1,02	

CAPO DI BUONA SPERANZA (1911 = 1)

1911 . .	31.064	226.463	1.428	34.143	1,00	1,00	1,00	1,00	
1918 . .	28.034	196.622	1.274	34.567	0,90	0,87	0,89	1,01	
1921 . .	28.042	199.845	1.303	34.669	0,90	0,88	0,91	1,02	
1926 . .	26.235	186.588	1.238	34.935	0,84	0,82	0,87	1,02	
1936 . .	23.079	157.383	1.045	34.953	0,74	0,69	0,73	1,02	

NATAL (1911 = 1)

1911 . .	28.308	192.329	1.247	34.579	1,00	1,00	1,00	1,00	1911
1918 . .	23.491	149.652	1.006	35.128	0,83	0,78	0,81	1,02	1920—21—22
1921 . .	24.165	155.704	1.036	34.979	0,85	0,81	0,83	1,01	1925—26—27
1926 . .	22.091	140.939	0.950	35.169	0,78	0,73	0,76	1,02	
1936 . .	18.915	118.633	0.791	35.016	0,67	0,62	0,63	1,01	

ANNI	VALORI EFFETTIVI				NUMERI INDICI				MEDIA DEI NATI NEGLI ANNI
	Natalità	Saggio fecond. legitt.	Indice K	Età media delle coniug.	Natalità	Saggio fecond. legitt.	Indice K	Età media delle coniug.	

CECOSLOVACCHIA (1921 = 1)

1921 . .	28.18	180.84	1.212	35.076	1,00	1,00	1,00	1,00	1920—21—22
1930 . .	21.83	126.04	0.814	34.520	0,77	0,69	0,67	0,98	1930—31—32

SERBIA (1900 = 1)

1895 . .	42.61	244.31	1.325	31.559	1,00	1,00	1,00	1,00	1895—96—97
1900 . .	39.43	222.44	1.201	31.473	1,08	1,10	1,10	1,00	1900—01—02

Esponiamo i risultati ottenuti per alcuni dei principali paesi europei.

In *Finlandia* nel 1950 i numeri indici risultano uguali a 0,63 per il coefficiente di Tait, 0,62 per la fecondità legittima e 0,76 per la natalità. Dal 1900 al 1950 l'indice di Tait è diminuito del 37 %, il saggio di fecondità legittima del 38 % e la natalità del 24 %. È da notare che il 1950 rispetto al 1940 segna un lieve aumento, più accentuato per la natalità. Nel 1880 i numeri indici sono 1,04 per il coefficiente *K*, 1,05 per la fecondità legittima e 1,12 per la natalità. L'età media in tutto il periodo di osservazione si mantiene pressochè costante.

In *Svezia* l'andamento decrescente della fecondità e della natalità è ancora più accentuato. Nel 1880 i numeri indici sono uguali a 1,07 per il coefficiente *K*, 1,06 per la fecondità legittima e 1,09 per la natalità; nel 1950 si hanno valori uguali a 0,41 per *K*, 0,42 per la fecondità legittima e 0,61 per la natalità. Dal 1900 al 1950 si verificano variazioni del 59 % (coefficiente *K*), del 58 % (fecondità legittima) e del 39 % (natalità). L'età media decresce lievemente: il numero indice del 1950 è 0,99.

In *Norvegia* nel 1950 i numeri indici risultano 0,49 (coefficiente *K*) 0,48 (fecondità legittima) e 0,64 (natalità) e denotano variazioni rispettivamente del 51 %, del 52 %, e del 36 %.

Nel 1855 l'indice relativo al coefficiente *K* è 1,04, alla fecondità legittima 1,07 ed alla natalità 1,11. L'età media può considerarsi pressochè costante.

Dal 1900 al 1950 in *Danimarca* l'indice di Tait decresce del 48 %, il saggio di fecondità legittima del 52 % e la natalità del 37 %; i numeri indici nel 1950 sono 0,52 (coefficiente K) 0,48 (saggio di fecondità legittima) e 0,63 (natalità). Variazioni accentuate si verificano anche per gli anni anteriori al 1900: nel 1850 si hanno valori uguali a 1,23 per il coefficiente di Tait, 1,06 per la fecondità legittima e 1,07 per la natalità. L'età media aumenta più che nelle altre nazioni (nel 1950 il numero indice è 1,04).

I numeri indici calcolati per l'*Inghilterra* e *Galles* risultano uguali a 0,48 per l'indice di Tait, 0,45 per la fecondità legittima e 0,55 per la natalità; le variazioni sono rispettivamente del 52 %, del 55 %, e del 45 %. Gli indici relativi al 1851 sono 1,16 (coefficiente K), 1,21 (fecondità legittima), 1,19 (natalità). Il numero indice relativo all'età media è nel 1950 uguale a 1,03.

L'*Irlanda* in tutto il periodo di osservazione presenta delle lievi variazioni. I numeri indici calcolati nel 1946 sono uguali a 0,95 (coefficiente K), 0,89 (fecondità legittima) e 0,98 (natalità). Nel 1871 si hanno indici uguali a 1,04 (coefficiente K), 1,99 (fecondità legittima), e 1,22 (natalità). L'età media nel 1946 è uguale a 1,02.

Le variazioni dal 1900 al 1947 della fecondità in *Belgio* sono presso a poco analoghe a quelle che si verificano in *Danimarca* dal 1900 al 1950. I numeri indici relativi al 1947 sono 0,52 per il coefficiente K , 0,48 per la fecondità legittima e 0,62 per la natalità. L'indice dell'età media è uguale nel 1947 a 1,03. Per gli anni anteriori al 1900 l'indice di Tait decresce in misura maggiore della fecondità legittima e della natalità. I numeri indici nel 1846 sono 1,32 (coefficiente K), 1,15 (fecondità legittima) e 1,00 (natalità).

Per i *Paesi Bassi* nel 1947 l'indice di Tait è uguale a 0,70, la fecondità legittima a 0,69 e la natalità a 0,87. Le variazioni sono rispettivamente del 30 %, del 31 % e del 13 %. L'età media si mantiene costante. Per il 1849 i numeri indici calcolati sono 1,17 (coefficiente K), 1,10 (fecondità legittima) e 1,07 (natalità).

In *Germania* i numeri indici risultano, nel 1950 0,41 per K ; 0,37 per la fecondità legittima e 0,47 per la natalità; l'età media è uguale a 1,04. Le variazioni sono tra le più forti delle

nazioni prese in esame. Nel 1871 i numeri indici sono 1,14 (coefficiente K), 1,11 (fecondità legittima e natalità).

Anche in *Austria* la fecondità e la natalità dal 1900 al 1950 decrescono fortemente. Gli indici del 1951 sono 0,39 per il coefficiente K , 0,35 per la fecondità legittima e 0,40 per la natalità. L'indice relativo all'età media è, nel 1951, 1,05.

La *Svizzera* ha nel 1950 numeri indici uguali a 0,49 (coefficiente K), 0,48 (fecondità legittima) e 0,64 (natalità), e le variazioni risultano rispettivamente del 51 %, del 52 % e del 36 %, analoghe a quelle della Norvegia. L'indice relativo all'età media dopo alcune oscillazioni è nel 1950 uguale a 1,00.

In *Francia* nel 1946 si hanno numeri indici uguali a 1,01 per il coefficiente K , 0,92 per la fecondità legittima e 0,91 per la natalità. È da notare che il 1946 segna una notevole ripresa della fecondità e della natalità rispetto agli anni precedenti. Il numero indice relativo all'età media è uguale a 1,03. Nel 1851 i numeri indici risultano 1,43 (coefficiente K), 1,35 (fecondità legittima) e 1,24 (natalità).

Nel 1940 in *Spagna* i numeri indici del coefficiente di Tait (0,71) e della fecondità legittima (0,67) mostrano variazioni rispetto al 1900 presso a poco analoghe a quelle che si verificano nei Paesi Bassi dal 1900 al 1947. Il numero indice relativo alla natalità, pari a 0,67, è invece notevolmente più basso. L'età media è in aumento (1,02). Nel 1887 si hanno numeri indici uguali a 1,00 per il coefficiente K , 1,01 per la fecondità legittima e 1,06 per la natalità.

In *Portogallo* nel 1950 l'indice di Tait risulta 0,69, il saggio di fecondità legittima 0,68 e la natalità 0,80. Le variazioni sono rispettivamente del 31 %, del 32 % e del 20 %, e non si discostano molto da quelle della Spagna. L'età media è uguale a 1,01. Nel 1864 si hanno numeri indici uguali a 1,07 per il coefficiente di Tait, 1,00 per la fecondità legittima e 1,05 per la natalità.

L'Italia è tra i paesi che presentano variazioni più lievi. Nel 1951 i numeri indici sono per il coefficiente K 0,77, per la fecondità legittima 0,66 e per la natalità 0,55. L'indice relativo all'età media è 1,06.

Nei paesi europei le variazioni meno rilevanti si verificano, per la fecondità, oltre che in Italia, in Irlanda, Spagna, Paesi

Bassi, Portogallo e Finlandia e per la natalità in Irlanda, Paesi Bassi, Portogallo, Finlandia, Norvegia e Svizzera.

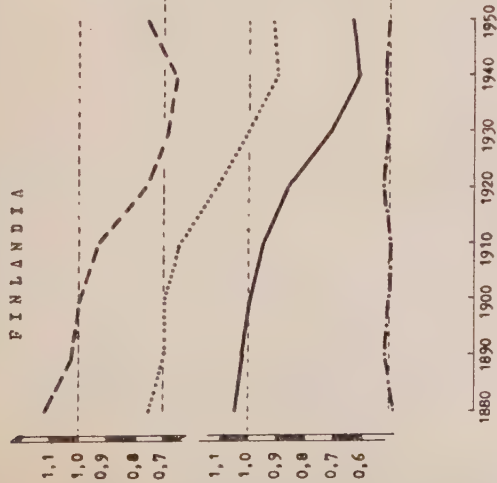
Per gli altri paesi europei ed extraeuropei l'esiguità di dati a disposizione non consente osservazioni significative.

In sintesi possiamo dire che, posto che la diminuzione di natalità, fecondità e indice K è comune a tutti i paesi e che l'età media è costante o in lieve aumento, l'andamento del saggio di fecondità legittima e quello dell'indice di Tait sono analoghi e la tendenza decrescente è lievemente più accentuata per il primo rispetto al secondo dal 1900 in poi, mentre meno rilevante appare la diminuzione della natalità.

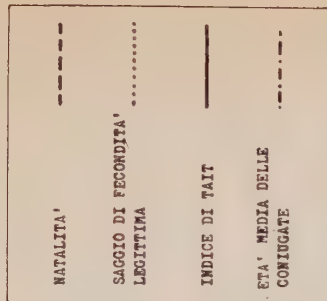
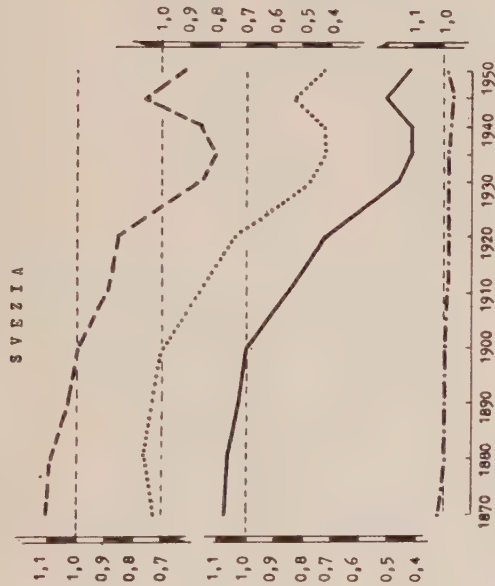
Questo si rileva in particolare per alcuni paesi come la Finlandia, dove la diminuzione dal 1900 al 1950 dell'indice di Tait è del 37 % e quella del saggio di fecondità legittima del 38 %, la Svezia dove la diminuzione dei saggi è sempre per lo stesso periodo rispettivamente del 59 % e del 58 %, la Norvegia dove si hanno variazioni decrescenti del 51 % e del 52 %, l'Inghilterra e Galles del 52 % e del 55 %, l'Olanda del 30 % e del 31 %, la Svizzera del 51 % e del 52 %, il Portogallo del 31 % e del 32 %.

Può concludersi dunque che l'indice di Tait è un indice sintetico la cui validità come misura della fecondità legittima è analoga a quella del saggio di fecondità legittima; l'esame dei grafici riportati in appendice, in cui sono rappresentati indice K , saggio di fecondità legittima e natalità ne dà conferma.

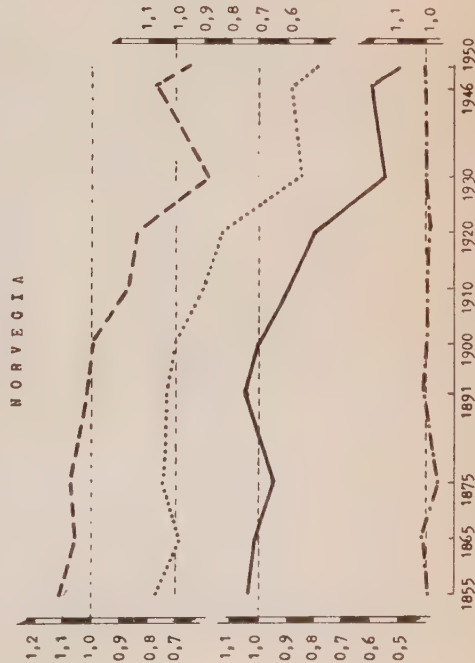
FINLANDIA



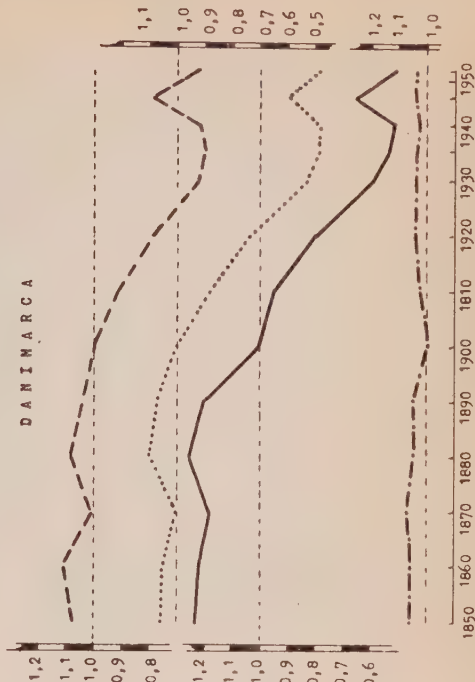
SVEZIA



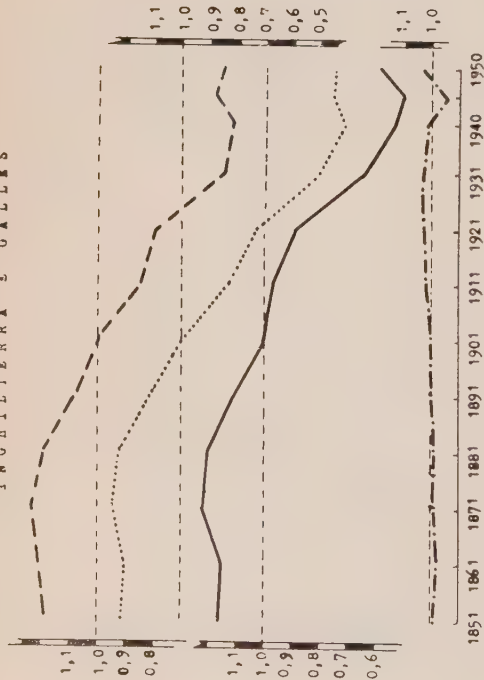
NORVEGIA



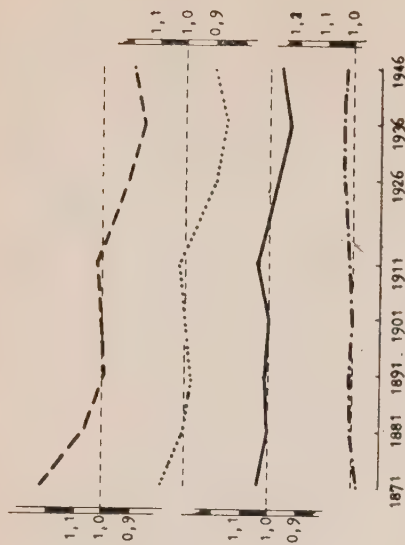
DANIMARCA



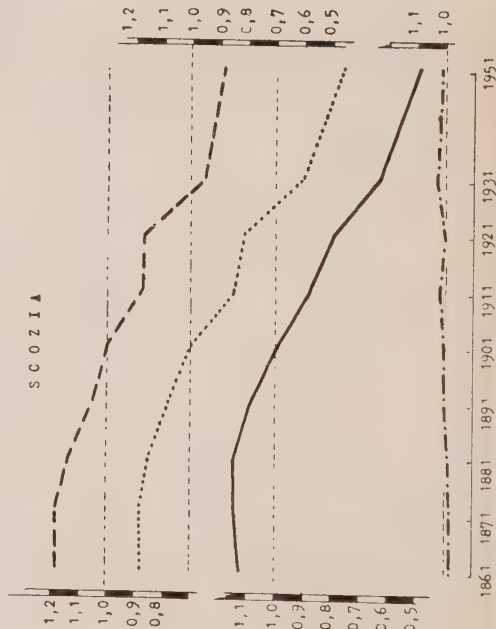
INGHILTERRA E GALLES



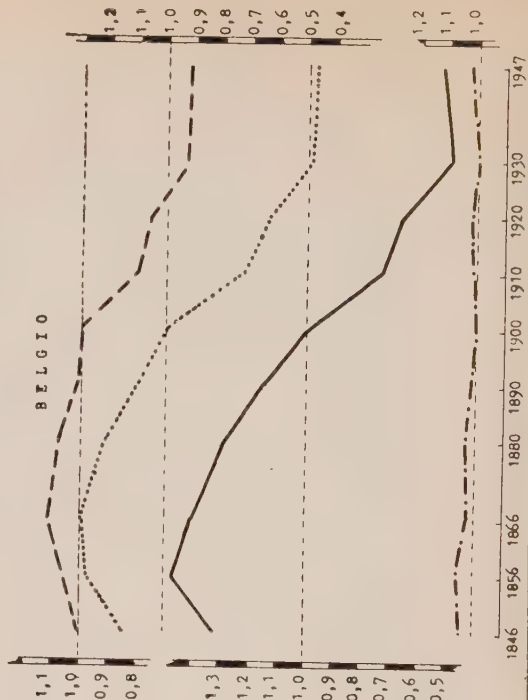
IRLANDA (EIRE)



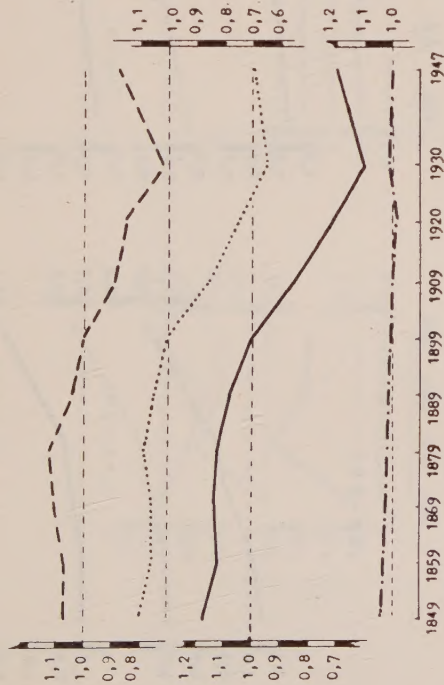
SCOTZIA



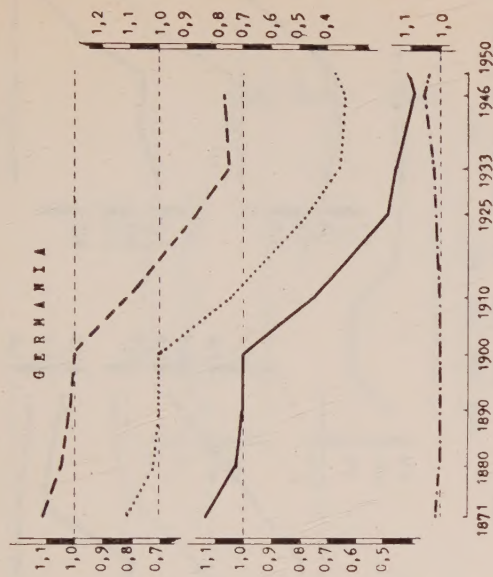
BELGIO



OLANDA



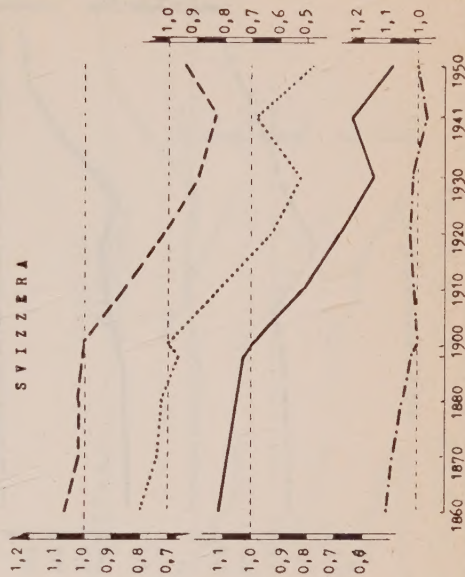
GERMANIA



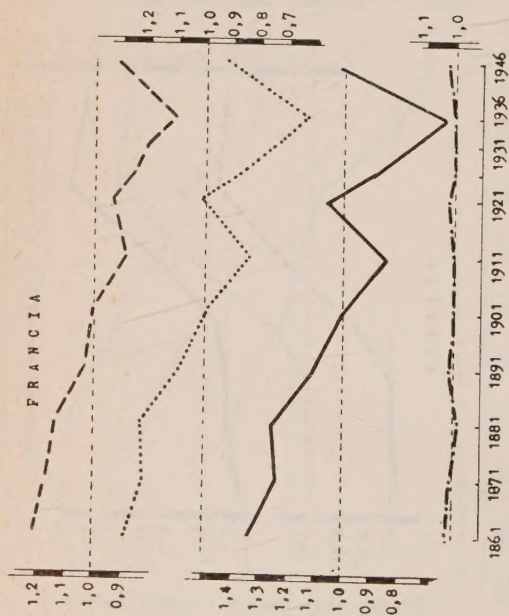
AUSTRIA



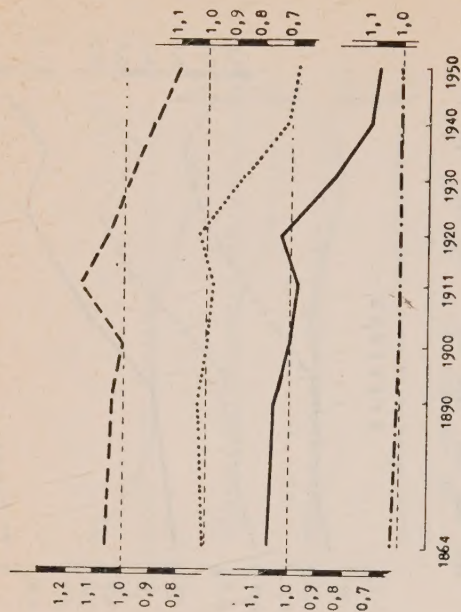
SVIZZERA



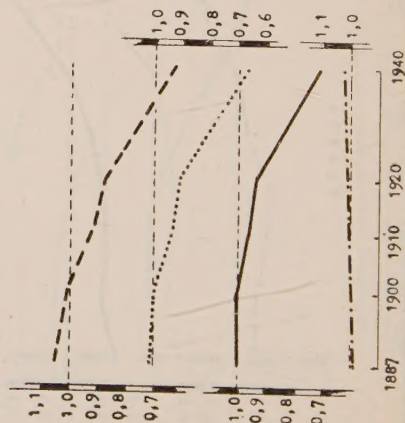
FRANCIA



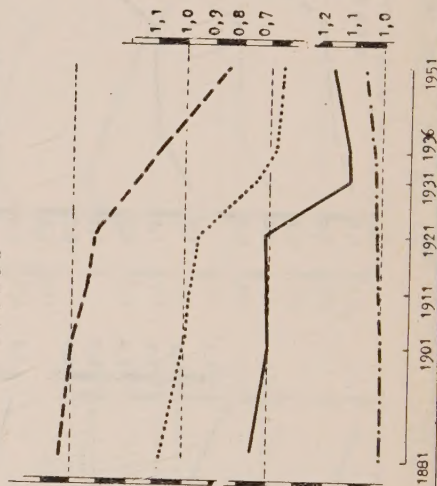
PORTOGALLO



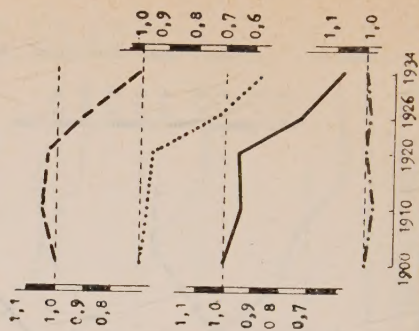
SPAGNA



ITALIA



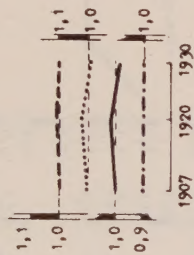
BULGARIA



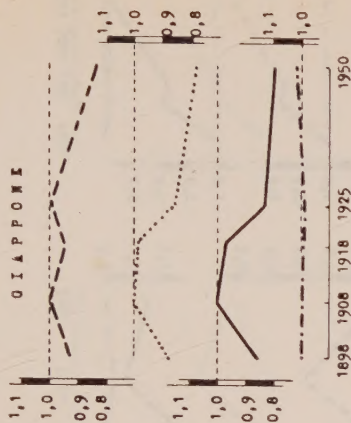
CANADA



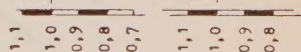
CILE



GIAPPONE



UNIONE SUD AFRICA



LIBERO STATO D'ORANGE



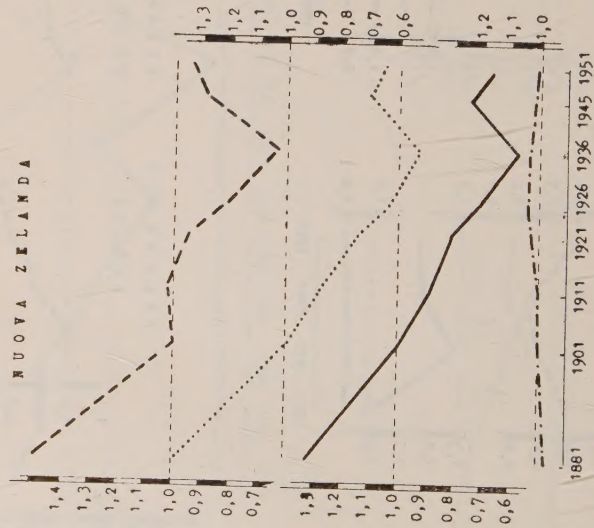
TRANSVAAL



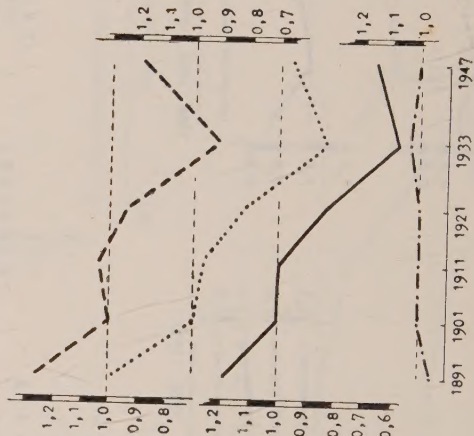
CAPO DI BUONA SPERANZA



NUOVA ZELANDA



AUSTRALIA



NATAL

